

SHARP OR BLUR: A FAST NO-REFERENCE QUALITY METRIC FOR REALISTIC PHOTOS

Fan Zhang

Lenovo Research, Shenzhen, China

zhangfanhk@lenovo.com

ABSTRACT

There is an increasing demand on identifying the sharp and the blur photos from a burst of series or a mass of collection. Subjective assessment on image blurriness takes account of not only pixel variation but also the region of interest and the scene type. It makes measuring image sharpness in line with visual perception very challenging. In this paper, we devise a no-reference image sharpness metric, which combines a set of gradient-based features adept in estimating Gaussian blur, out-of-focus blur and motion blur respectively. We propose a dataset-adaptive logistic regression to build the metric upon multiple datasets, where over half of the samples are realistic blurry photos. Cross validation confirms that our metric outperforms the state-of-the-art methods on the datasets with a total of 1577 images. Moreover, our metric is very fast, suitable for parallelization, and has the potential of running on mobile or embedded devices.

KEYWORDS

Image sharpness, No reference metric, out-of-focus, motion blur, logistic regression

1. INTRODUCTION

With fast-growing consumer electronics camera technology, such as phone camera, wearable camera, vehicle camera and aerial camera, there are challenging times in exploring easily attained photos. Those photos might be captured in a causal way, without a stable support to camera or intent focusing on scene. One demand is to discard the useless blurry photos in a mass of collection for efficient album management. Another challenge is to pick up the clearest photo from a sequence of burst shooting, for either photo snap under unavoidable camera shaking or photo recognition crowd-sourcing service like *CamFind*. To address the automatic and instant photo selection, there is a growing interest in a photo sharpness (or blurriness) metric in line with human visual perception.

Blur analysis has been widely studied but not well solved. Blur identification in computer vision society aims at estimating the type and the amount of blur [19, 16, 28, 26]. It tells in-focus regions from out-of-focus ones or moving regions from still backgrounds, forming a blurriness map to guide segmentation [5], super-resolution [9], shape-from-focus [20], depth-from-defocus [32], motion-from-blur [11], defocus magnification [1], and deblur [11, 25]. However blur identification pays attention to local blurriness in radiometric intensity rather than the overall quality in perception. Focus measure in electronic microscope and camera design [13, 20] can

tackle a focus image series, but probably fail when the scene is moving or changing. Quality metrics SSIM, widely used in signal processing society, match perception fairly well, but they need a reference image, known as the full-reference metrics. Recent studies develop the no-reference sharpness metric [14, 2] with the guidance of subjective image quality database and even unlabeled images. The subjective databases provide the MOS (subjective mean opinion scores) as the ground truth of blur extent, and the unlabeled data are synthesized with the same distortion type and level as the subjective databases [29, 27]. However, most databases use the synthetic images, which are generated from a limited set of source images with spatially-invariant Gaussian convolution. Existing metrics often do well in those synthetic datasets,

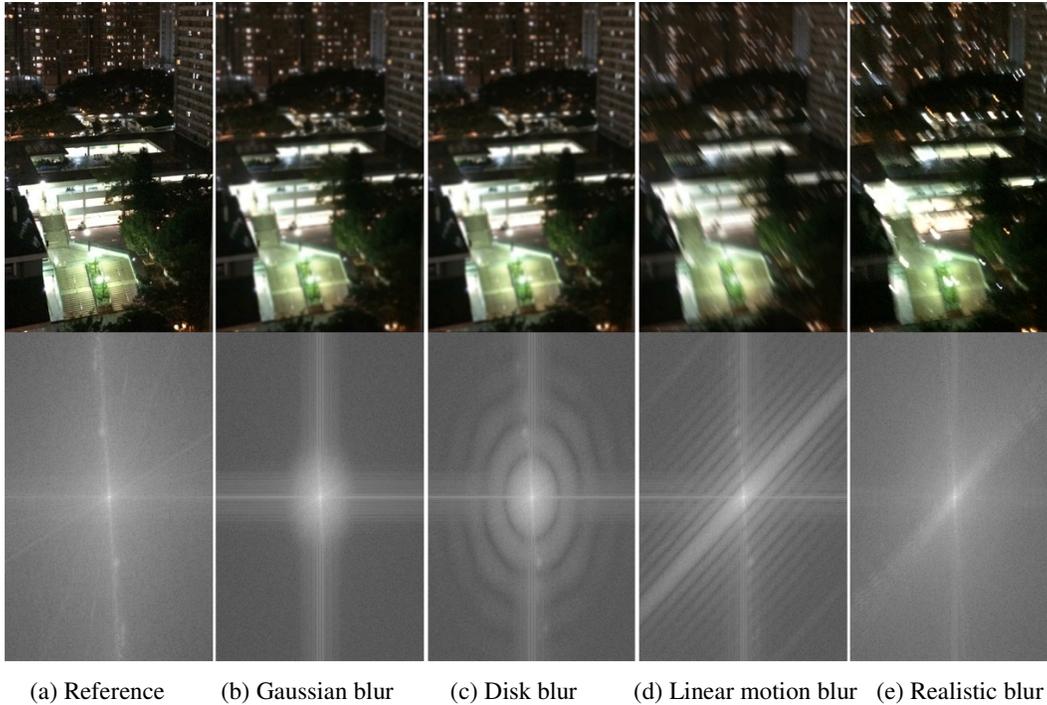


Figure 1. Synthetic and realistic blurry images and their log spectrum $\log|\mathcal{F}(f)|$

but perform poorly on the realistic blur, e.g., UFRJ database [10, 7]. It is also interesting that Ye et al. studied the relationship between the blurriness of document images and the OCR (optical character recognition) accuracy, and reported that OCR accuracy may not be consistent with human perception [29]. To summarize, few prior arts study the realistic blur from the perspective of perception; the existing metrics are good at synthetic blur among an image series sharing the similar scenes, but perform poorly in practice and change dramatically when scene changes.

In fact, realistic blur is challenging to measure because:

1) Realistic blur mainly include out-of-focus blur and motion blur.

Out-of-focus blur generally smooth the edges, but motion blur may generate edges parallel to the motion smear, e.g., light streaks (see Fig. 1(e)). A single feature can hardly predict the hybrid blur.

2) Photos may exhibit diverse scenes, from the smooth, e.g., sky and face, to the rough, e.g., forest and fabrics.

A natural scene may contain steep edges around occlusions or soft edges due to illumination. Simple image statistics often fail to tell a sharp smooth scene from a blur rough scene.

3) Blur pixels do not always degrade photo sharpness.

For example, lens blur can pop out the in-focus objects from the out-of-focus background and make objects distinctively sharper in appearance. Pure average of local sharpness is possibly inconsistent with subjective appreciation.

In this study, we develop a fast metric for practical applications. Keeping this in mind, we learn from realistic data and approximate the solution by using a set of low-level vision features. Contributions of our work include:

1) We design and select a set of features regarding their correlations with various perceptual blur. Those features employ different image statistics and varied pooling strategies to complementarily measure the multiple types of blur in subjective datasets.

2) Such datasets are an exhaustive gathering of current public subjective databases and our own collection of failure cases. Those versatile data may guarantee unbiased data-driven modeling.

3) We formulate a dataset-adaptive logistic regression to co-train our metric on multiple datasets. It bridges the gap between misaligned datasets and makes full use of the adopted data-driven approach.

2. REALISTIC BLUR

In photography, photo blur mainly stems from lens blur and motion blur. Lens blur keeps clear the in-focus objects if any, and yet obscures the out-of-focus things. Motion blur may pervade the whole photo for camera shake or occurs locally on moving objects.

In mathematics, out-of-focus blur is modeled as convolution with a disk kernel, and motion blur as convolution with a trajectory kernel. If the motion happens to be linear uniform during exposure, the kernel evolves to a line. A disk kernel has a spectrum with circularly symmetric sinc waves, while a line kernel exhibits parallel sinc waves in spectrum [12, 16, 28]. If a kernel is **spatially-invariant**, its distinct pattern accumulates and reflects in the image spectrum. Fig. 1 shows the log spectrum of the disk blur (c) and linear motion blur (d) synthesized from the reference (a).

However, the spatially-invariant assumption is too strong for realistic blur. Many factors will violate the assumption.

1) Out-of-focus varies with the object depth, and the motion blur correlates to the velocity of moving objects;

2) Lens geometric distortion differentiates the blur along the radial direction;

3) Nonlinear tone mapping (e.g., Gama correction) in imaging pipeline changes the dependency among pixels which originally takes place at image sensor.

Moreover, after the processing in imaging pipeline including demosaic, denoise and/or resize, the final blur makes toward Gaussian blur to some extent. Fig. 1(e) shows the log spectrum of

realistic blurry image, where sinc wave patterns disappear. As a result, realistic blur is often hybrid and hardly described by a simple mathematic model.

3. PERCEPTUAL SHARPNESS FEATURE

A considerable amount of subjective rating data is available now. It inspires us to use data driven modeling. First of all, we design and select the basic features for each type of blur, including Gaussian blur, out-of-focus blur and motion blur. Then we combine the features with logistic model and train a robust metric on datasets with regression method.

3.1. Maximal gradient (MAG)

Image gradient can indicate the image sharpness, for blur smooths the gradients. For example, Gaussian blurring process is the solution to the diffusion equation of $\frac{\partial I}{\partial t} = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}$ in physics. The image intensity can be considered as a flow, that is pushed away by a force equal to the gradient. The steeper the gradient, the faster the pixels diffuse. Considering the directional gradients from the current pixel to its 8-connected neighbors as denoted in Fig. 2, the steepest gradient is defined as $\max_{d=0,1,\dots,7} |I_d(x,y)|$ by Bahrami [2].

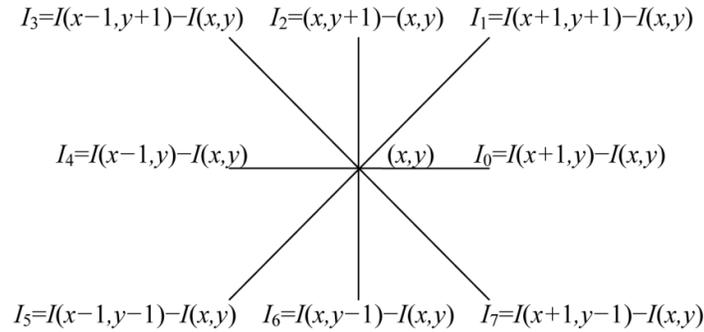


Figure 2. Notations of directional gradients I_d at pixel (x,y)

Smooth regions always pervade natural images, where steep gradients rather than gentle gradients dominate the perceptual sharpness. So the largest local maximal gradients are taken into account, which often correspond to the structures and textures of an image. As proposed in [2], the global maximal gradient is computed as

$$\text{MAG} = \mathbf{E} \left[\max_k \left\{ \max_{(x,y)} \left\{ \max_{d=0,1,\dots,7} |I_d(x,y)| \right\} \right\} \right], \quad (1)$$

where operator $\mathbf{E}[\cdot]$ computes the expectation (i.e., the mean), k is empirically set to 2% and thus the operator $\max k$ takes 2% of the largest local maximal gradients. The local maximal gradient maps are shown in the second row of Fig. 3, where 2% of the largest values are marked bright.

3.2. Minimal 2nd order gradients ratio (MGR)

Gaussian blur and out-of-focus blur smooth image isotropically, while motion blur flattens image specially along the motion smears. Such flattening can decrease the 2nd order image gradient in a certain direction. Blurriness in this direction can be estimated by the minimal 2nd order gradient, i.e., $\min_{d=0,1,2,3} |I_d(x,y) + I_{d+4}(x,y)|$.

However, smooth regions also have small 2nd order gradients, and thus may mislead the blurriness measurement above. To overcome the flaw, we, again, select and count the largest values within an image like MAG, as

$$\text{M2G} = \mathbf{E} \left[\max_{(x,y)} k \left\{ \min_{d=0,1,2,3} |I_d(x,y) + I_{d+4}(x,y)| \right\} \right]$$

The minimal 2nd order gradient maps are shown in the third row of Fig. 3, where the largest values are marked bright.

A heavier motion blur can spread over a wider area, but can be hardly described by the 2nd order gradient which covers only a 3×3 neighbourhoods. To avoid slow convolutions with big kernel size, we resort to multi-scale analysis. To be specific, the ratio of the M2G values between image scales may indicate whether the blurriness propagation terminates at the current stage. Given a

sequence of increasing resolutions $(I_{2^j})_{j \in \mathbb{Z}^+}$ we have corresponding M2G_{2^j} :

$$\mathbf{E} \left[\max_{(2^j x, 2^j y)} k \left\{ \min_{d=0,1,2,3} |I_d(2^j x, 2^j y) + I_{d+4}(2^j x, 2^j y)| \right\} \right]$$

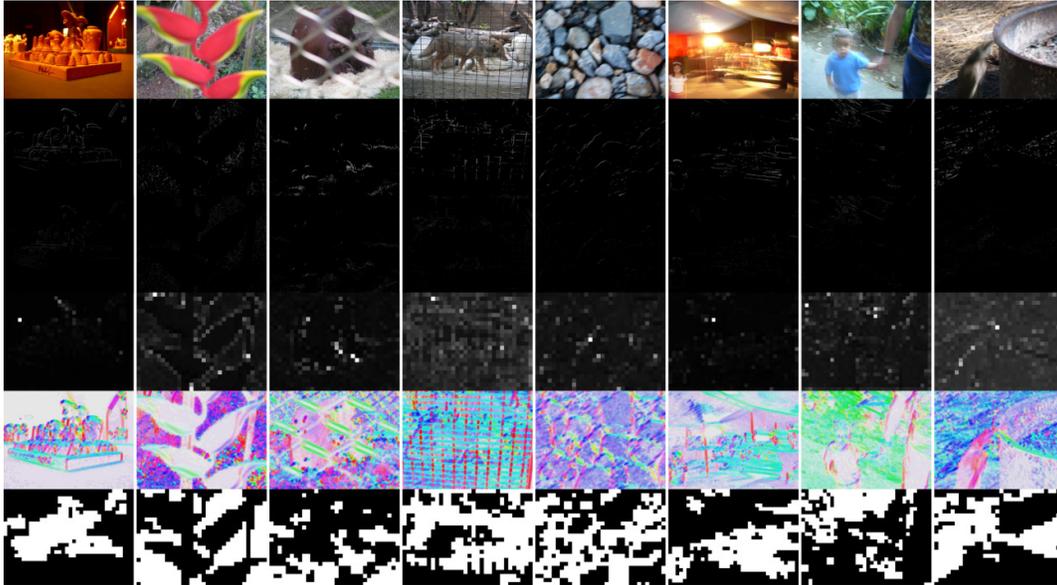


Figure 3. Visualized sharpness feature map of the samples in UFRJ database. The rows, top down, show original images, maximal gradients, minimal 2nd order gradients, gradient kurtosis, doubled gradient angles, and patch-wise MAG (for computing MSG) respectively.

We define the minimal 2nd order gradients ratio (MGR) as:

$$\text{MGR}_1 = \text{M2G}_{2^1} / \text{M2G}_{2^0} \quad (2)$$

$$\text{MGR}_2 = \text{M2G}_{2^2} / \text{M2G}_{2^1} \quad (3)$$

3.3. Average gradient kurtosis (AGK)

Natural images have super-Gaussian distribution with an acute peak and heavy tails in spatial frequency domain. The blur process widens the distribution and thus reduces its peakedness. Actually, the kurtosis for a blurred patch is smaller than that of a sharp one in Fourier [31], DCT (discrete cosine transform) [4], and gradient domain respectively [25]. Shi et al., use local gradient kurtosis to justify the blurriness of patches [25], which inspires us evaluating the sharpness of whole image with the average local gradient kurtosis as

$$\text{GK} = \mathbf{E}_{p \in I} \left[\min\{K_p|I_0, K_p|I_2\} \right] \quad (4)$$

$$\forall p \in I, \forall (x, y) \in p : K_p = \frac{\mathbf{E} \left[(I(x, y) - E[I_p])^4 \right]}{\mathbf{E} \left[(I(x, y) - E[I_p])^2 \right]^2}$$

where image grid I is divided into non-overlapped patches indexed with p , kurtosis operator $K_p[\cdot]$ computes the kurtosis within the p -th patch, and the local gradient kurtosis is the smaller one between the vertical gradient kurtosis and the horizontal gradient kurtosis. The gradient kurtosis maps are shown in the fourth row of Fig. 3.

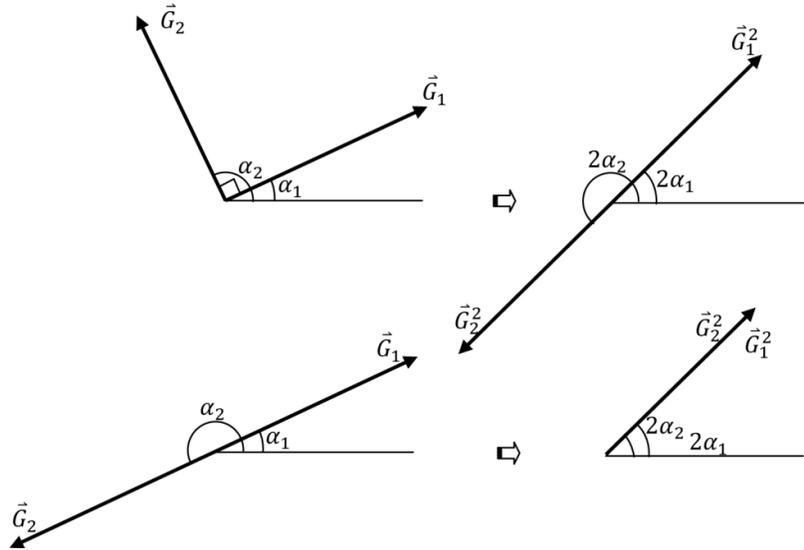


Figure 4. Doubling the vector angle may cancel out the perpendicular vectors (top) and consolidate the opposite vectors (bottom).

3.4. Average angle-doubled gradient

So far, we expect that the maximal gradients can measure the isotropic blur like Gaussian blur and the minimal 2nd order gradients can evaluate the anisotropic blur like simple motion blur. Though blur classification is simplistic in practice, we nevertheless try to identify the blur type. Note that linear motion blur smooths image along and opposite to the motion direction and meanwhile maintain or even enhance the contrast perpendicular to the motion direction. Measuring the coherence of gradient angles should therefore add the same or the opposite directions but cancel out the orthogonal directions. As proposed by Jang et al. [15], doubling and subsequently averaging the gradient vectors can exactly accomplish this goal. Actually as proved in [8], the average doubled gradient angle correlates to the direction, along which the image embodies the minimal high-frequency energy.

The gradient vector field ∇I can be represented by the complex array $I_0 + iI_2$, and doubling arguments (angles in radius) is obtained by squaring the vectors as $\nabla I^2 = I_0^2 - I_2^2 + i2I_0I_2$. Adding the angle-doubled gradients may cancel out the perpendicular gradients and yet consolidate the opposite gradients, as illustrated in Fig. 4. In Fig. 3, the angle-doubled gradient maps are visualized with uniform-lightness color images, with the double angle as hue and the magnitude as saturation. Coherent motion directions leads to concentrated colors. On one hand the energy of average angle-doubled gradient (EDG) indicates the coherence.

$$|\overline{\nabla I^2}| = \mathbf{E}[I_0(x, y)^2 - I_2(x, y)^2]^2 + \mathbf{E}[2I_0(x, y)I_2(x, y)]^2$$

The bigger the magnitude is, the more directionally coherent the gradients are. On the other hand, the average energy of angle-doubled gradients (ADG), as

$$\begin{aligned} |\overline{\nabla I^2}| &= \mathbf{E}\left[(I_0(x, y)^2 - I_2(x, y)^2)^2 + (2I_0(x, y)I_2(x, y))^2\right] \\ &= \mathbf{E}[(I_0(x, y)^2 + I_2(x, y)^2)^2] \end{aligned} \quad (5)$$

equals to the average squared gradient energy, and reflects the contrast of textures and edges in an image. Chen et al. argue that ADG can identify whether the image (patch) is motion blurred or out-of-focus blurred [8], while Jiang et al. suggest that EDG normalized by ADG can identify blur type more inconsistently with scene change [15]. We on one hand select ADG as a feature of our metric, and on the other hand normalize MDG by ADG to obtain the normalized energy of average angle-doubled gradient (NDG):

$$|\widetilde{\nabla I^2}| = |\overline{\nabla I^2}| / |\overline{\nabla I^2}|$$

To identify heavier blur, we use multi-scale analysis again and define the product of NDGs across image scales as

$$\text{PNDG} = |\widetilde{\nabla I^2}|_{2^0} |\widetilde{\nabla I^2}|_{2^1} |\widetilde{\nabla I^2}|_{2^2} |\widetilde{\nabla I^2}|_{2^3} \quad (6)$$

3.5. Moment of sharp gradients (MSG)

How a local patch contributes to perceptual global blurriness may depend on where it is. For example, an in-focus subject against out-of-focus background, which is often created in a shallow DOF (depth of field), will look distinctively sharper. Therefore, the distribution of sharp regions and blur regions should be taken into account.

First of all, we classify an image into a shallow DOF or a deep DOF by analyzing the concentration of sharp regions. To this end, we estimate a patch-wise binary sharpness map for an image. During computing the aforementioned MAG, we have located the k -largest gradients in an image. A patch that contains more than T top- k -largest gradients belongs to the sharp patch set, denoted by $\mathcal{P} \in \mathcal{P}_s$, otherwise the blur patch set, denoted by $\mathcal{P} \in \mathcal{P}_b$. As a result, concentrated large gradients will cause fewer sharp patches while scattered large gradients will generate more sharp patches. The resultant gradient maps are shown in the bottom row of Fig. 3, where sharp patches are marked bright and blur patches are assigned dark. Accordingly, an image with less than 40% sharp patches is classified as a shallow DOF scene, denoted by $I \in \mathcal{S}_s$ otherwise as a deep DOF scene, denoted by $I \in \mathcal{S}_d$.

Then, we analyse the distribution of sharp and blur regions. In photography, the subjects are typically framed and composed at the center or at the one third in the image space (known as rule of thirds), where we mark as *composition reference points*. We define *moment arm* as the city block distance from the considered point (x, y) to the nearest composition reference point, as

$$a(x, y) = \min \left\{ \left| x - \frac{W}{3} \right|, \left| x - \frac{W}{2} \right|, \left| x - \frac{2W}{3} \right| \right\} + \min \left\{ \left| y - \frac{H}{3} \right|, \left| y - \frac{H}{2} \right|, \left| y - \frac{2H}{3} \right| \right\}$$

where W and H are the width and height of the image. We further define the sharp moments as the average transformed moment arms for all sharp points,

$$\forall (x, y) \in P_s : m_s = \mathbf{E} \left[\frac{1}{1 + r \cdot a(x, y)} \right]$$

and the blur moments as the average transformed moment arms for all blur points.

$$\forall (x, y) \in P_b : m_b = \mathbf{E} \left[\frac{1}{1 + r \cdot a(x, y)} \right]$$

Finally, we calculate the feature for shallow DOF and deep DOF scene separately:

$$\text{MSG} = \begin{cases} m_s & \text{if } l \in S_s \\ 1 - m_b & \text{if } l \in S_d \end{cases} \quad (7)$$

4. ADAPTIVE REGRESSION ON MULTIPLE DATASETS

Given an image set $\{I_i\}$ with the mean opinion scores about perceptual sharpness $\{MOS_i\}$ and a group of selective features $\{f\}$, we look for a metric $q(f; \beta)$ with parameters β , to maximize the likelihoods:

$$\max_{\beta} \sum_i \mathcal{L}(q_i(f; \beta); MOS_i)$$

If multiple image sets are available, we maximize the total likelihoods among all the sets:

$$\max_{\beta} \sum_j \sum_i \mathcal{L}(q_{i,j}(f; \beta); MOS_{i,j}) \quad (8)$$

where the datasets are indexed by j .

Subjective opinions about perceptual sharpness are always bounded, like other psychological measurements. That is, opinion scores will approach the lower bound for the worst quality and the upper bound the best quality, which are called the flooring and the ceiling effects.

A key challenge of analyzing multiple datasets is that the mean opinion scores cannot be mixed up straightforwardly. This is because the flooring and ceiling effects rely on the context of the test materials and may vary across datasets. Actually, a human subject can hardly value an isolated photo without being demonstrated how the “best” and the “worst” ones look like. The test conditions between the datasets were factually inconsistent here. So the transform from the sharpness measure towards the MOS cannot be fixed since the context is not consistent.

To span the gap, a standard of method is to align the MOSs via the “anchor” samples shared between datasets [3]. However, the datasets here cannot be aligned in that way, for lack of intersect images as anchors. We propose adaptive logistic regression for the problem as following. Logistic modeling is suitable for psychological measurement, since it fuses features and maps them to a bounded interval. A logistic model can be written as

$$q = \frac{1}{1 + \exp(\beta^T \cdot f + b)} \quad (9)$$

where β is a parameter vector with the equal length as feature vector f and b is a scalar parameter. Traditionally, b is often merged into β as β_0 . However, we isolate b and adapt it to each dataset. That is, we use a unique β for all datasets and yet $\{b_j\}$ for the datasets indexed with j . So Eq.(8) is rewritten as

$$\max_{\beta, b} \sum_j \sum_i \mathcal{L}(q_{i,j}(f; \beta, b_j); MOS_{i,j}) \quad (10)$$

We assume the opinion scores as Gaussian distributed variables and therefore instantiate the likelihood as mean squared error. Eq.(10) is derived as

$$\min_{\beta, b} \sum_j \sum_i \left[MOS_{i,j} - \frac{1}{1 + \exp(\beta^T \cdot f_{i,j} + b_j)} \right]^2 \quad (11)$$

Parameters b_j controls the shape of the sigmoid logistic curve and adapts the predicted quality scores q towards the MOSs for each dataset. They compensate the misaligned flooring and ceiling effects across datasets. It is interesting that b does not change the rank order of the predict quality scores for each dataset. Instead, it is β that determines the rank order of the predicted quality.

It is straightforward to use the nonlinear regression toolbox of MATLAB to solve β with a gradient-descent method. The convergence is usually guaranteed during our ample random tests. Moreover, the regression toolbox also provides the statistical significance test for the feature evaluation, which will be discussed in the next section.

5. EXPERIMENT

5.1. Protocol

We collect nine datasets, including the publicly-available databases and our own database. We extract the subsets of Gaussian blurred images, from the public databases LIVE [24], IVC [18], A57 [6], TID2008 [22], CSIQ [17], VCL_FER [30], and TID2013 [21]. These images are synthesized by convoluting sharp reference images with spatially-invariant Gaussian kernels. UFRJ database [10] and our database, contain realistic blurry photos. In UFRJ database, the photos are captured with digital compact cameras, and wherein the blur is further classified and labeled as out-of-focus, simple motion, complex motion and “other” type. Our database contains the photos snapped with phone cameras and wearable cameras in daily life. Its samples are the failure cases during our many rounds of redesign and retest (refer to the supplementary material for more detail). We believe that those data allow us avoid overfitting on limited and biased data. Both UFRJ and our database keep the JPEG EXIF information intact for each image.

The full datasets contain a total of 1577 blurry images and the associated MOSs. The MOSs represents the ground truth of image sharpness or blurriness, and is used to evaluate the prediction accuracy of metric. In experiment, we normalize all subjective scores to the range [0,1]; a MOS of 0 indicates the worst quality (the blurriest) while a MOS of 1 represents the best quality (the sharpest).

5.2. Evaluation criterion

The model accuracy is evaluated using the Spearman’s rank order correlation coefficient (SROCC) ρ_s between the predicted and the subjective quality score series. ρ_s has a range of $[-1, 1]$; the higher the value, the better the accuracy. Random predictions will achieve a ρ_s value of about 0.

ρ_s evaluates the ordinal match between two score series, and thus remains invariant with any a monotonic mapping of the series. In other words, the accuracy in terms of ρ_s does not rely on any curve-fitting procedure. Such a curve-fitting procedure is, however, inevitable and sensitive for computing other criteria, like Pearson’s linear correlation coefficient (LCC) and root-mean-squared error (RMSE).

5.3 Feature evaluation

The selective features f includes MAG, MGR₁, MGR₂, AGK, ADG, PNDG, MSG as well as EXP (exposure time), as listed in Table 1. The exposure time is recorded in the JPEG EXIF data. For the image without the exposure time information, we set EXP = 0.01 second (at such a shutter speed, the photo just tends not to blur). We compared the proposed features and state-of-the-art metrics over the datasets. The accuracy for each type of blur is plotted in the spider chart of Fig. 5, where the radial axis indicates the correlation ρ_s ranging from -0.4 at the center origin to 1 at the outermost square grid. Four types of blur are counted. For Gaussian blur (the upward direction in Fig. 5), we compute the average ρ_s on the seven Gaussian blur datasets, which cover a total of 687 images. For out-of-focus (rightward), simple motion (downward) and complex motion (leftward) blur, we calculate the ρ_s on the corresponding subsets of UFRJ database, which contain 141, 57 and 62 images respectively.

Table 1. Abbreviations and definitions of features and metrics.

Abbr. (Def.)	Feature / metric operator [Ref.]
ADG (5)	Average energy of angle-Doubled Gradients
AGK (4)	Average Gradient Kurtosis
EXP	Exposure time in JPEG EXIF
GRA1	Gaussian derivative [20]
GRA2	Gradient energy [20]
GRA4	Squared gradient [20]
LPC_SI	Local phase congruency sharpness index [14]
MAG (1)	Maximal Gradient
MGR ₁ (2)	Minimal 2 nd Gradient Ratio at Scale 1 & 2
MGR ₂ (3)	Minimal 2 nd Gradient Ratio at Scale 3 & 4
MIS9	Vollath’s autocorrelation [20]
MSG (7)	Moment of Sharp Gradients
PNDG (6)	Product of Normalized energy of average angle-Doubled Gradient
STA8	Histogram range [20]

Table 2. Statistical significance of features in regression

Feature f_m	95% CI of Parm. β_m	p value of Parm. β_m
EXP	0.290±0.042	4.0×10^{-39}
MGR ₂	-1.07±0.19	1.2×10^{-28}
MGR ₁	-0.363±0.099	1.0×10^{-12}
PNDG	0.0796±0.0277	1.0×10^{-11}
ADG	-0.346±0.142	2.1×10^{-6}
MAG	0.242±0.142	9.0×10^{-4}
AGK	-0.359±0.221	1.5×10^{-3}
MSG	-0.762±0.586	1.1×10^{-2}

These experimental results give us a first impression about the capability of the features. Among the proposed features, ADG has the best overall accuracy on all datasets; MGR₂ ranks the second in overall and is specially good at “simple motion” and “out-of-focus” blur; AGK is generally accurate and does especially well in motion blur; MAG achieves the state-of-the-art accuracy on

Gaussian blur; and the other feature MGR_1 , MSG or $PNDG$ alone appears not to correlate with each type of blur very well.

We compare with the state-of-the-art Gaussian blur metric LPC_SI [14], the best focus metric for microscope $MIS9$ [13], and the recommended focus metrics for OCR, i.e. $GRA1$, $GRA2$, $GRA4$, and $STA8$ [23] (see their definitions in [20]). Considering the performance of the existing metrics as shown in Fig. 5(b), the Gaussian blur is the easiest to predict, the out-of-focus blur is also easy to handle, and yet the motion blur are much more challenging to measure. This is partly because the Gaussian blur is synthetic and ideally spatially-invariant, the out-of-focus blur here is also nearly spatially-invariant since the images with shallow DOF have been picked out to the “other” type of blur in UFRJ database, but the motion blur here is rarely coincidental with spatially-variant blur. Another reason is that the Gaussian blur images, although abound here, are synthesized and derived from a few reference images, but the images in UFRJ database have more diverse scenes.

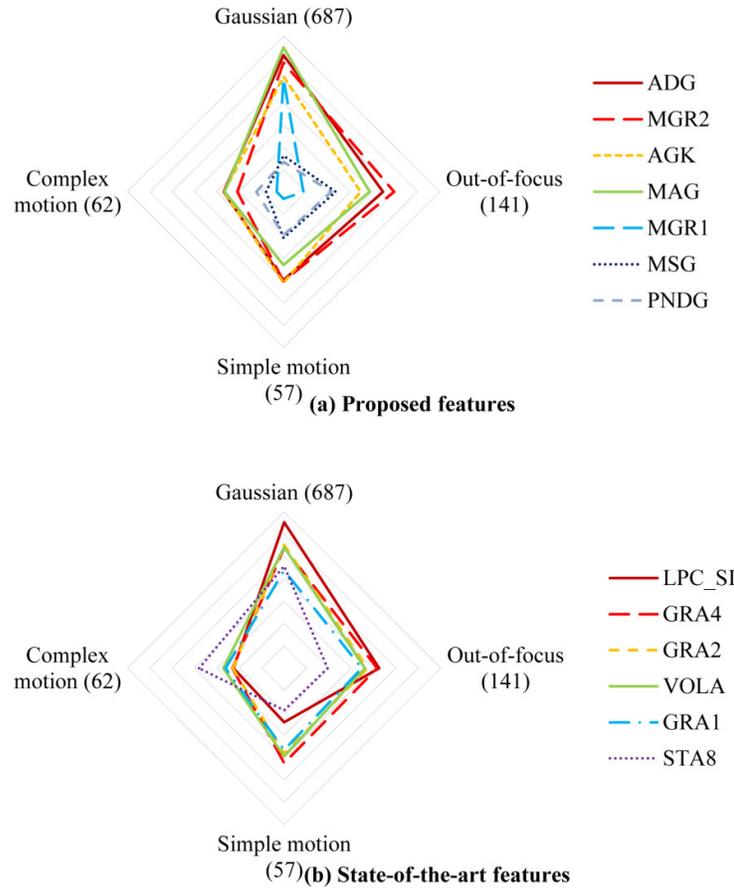


Figure 5. Correlation ρ_s between features and single type of perceptual blur, in terms of spider chart

Feature selection not only depends on the performance of using each feature alone, but also relies on that of using feature combinations. The latter point can be evaluated by the statistical significance test. For logistic regression model, the p value and the CI (confidence interval) can indicate the significance of the feature variables. On one hand, the smaller the p values is, the more confident the corresponding feature is. On the other hand, the 95% CI (confidence interval) of β_m does not cover 0 means that feature f_m is sufficiently confident. In this paper, we omit the process of feature selection but show the statistical significance of the final feature combination in Table 2. Parameters $\beta = \{\beta_m\}$ are obtained by nonlinear regression on the full datasets. All the

95% CIs are apart from 0, so it is confident that the selective features are powerful. It is worth to mention that MGR_1 and $PNDG$ correspond to quite small p values in Table 2, despite low ρ_s values in Fig. 5. It means that MGR_1 and $PNDG$ themselves alone are weak but they are really helpful to the feature combination. Moreover, the image composition related feature MSG also plays a fairly significant role, for a p value of 0.01.

5.4 Metric comparison

With the selective features and adaptive logistic regression, we obtain the metric as Eq. (9). The accuracy of metrics is compared in Fig. 6. In Fig. 6, the radial axis still indicates ρ_s on each dataset ranging from -0.4 at the centre to 1 at the outermost decagon grid. In the upward direction, we compute the weighted average ρ_s on all datasets, which is weighed by the number of images in each dataset.

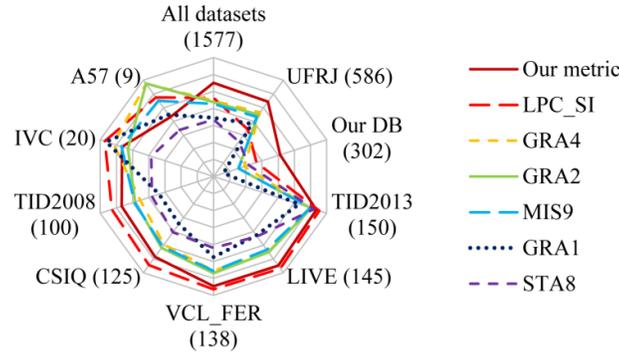


Figure 6. Correlation ρ_s of metrics on each dataset

For a fair comparison, we report the cross validation result of our metric. That is, for each time, we randomly divide each dataset into two segments, 50% for training and the other 50% for testing, find a unique set of parameters $\{\beta\}$ by training, and test it to obtain a set of accuracy on every dataset. Running the procedure for 100 times, we compute the average accuracy.

As a result, our metric achieves the best overall accuracy on all databases. It outperforms other metrics on the realistic photos datasets, UFRJ and our database. UFRJ database is quite challenging. To the best of our knowledge, only two papers disclose the result of their proposal on UFRJ database; the authors of UFRJ database report ρ_s of 0.56 ± 0.04 [10], and Chen et al. claim ρ_s of 0.586 [7]. Our metric attains ρ_s of 0.688 ± 0.052 in cross validation. We use the same features but replace logistic regression with SVR (supporting vector regression), and obtain a ρ_s of 0.631 in cross validation. Our database is even tougher, since it contains the failure cases during our past repetitive trials, such as blurry but high-contrast images (e.g., textural, noisy and overexposure scenes) as well as sharp but low-contrast scenes (e.g., sky, lake and nightscape), as shown in Fig. 7. On our database, LPC_SI only obtains ρ_s of 0.132 and most existing metrics even get a negative ρ_s , which are not better than a random prediction. However, our metric attains ρ_s of 0.424.

For most of the synthetic image datasets, our metric achieves comparable accuracy as the state-of-the-art approach LPC_SI . It is inferior to most metrics only on A57 and IVC. These two datasets are too small to guarantee unbiased random divisions in cross validations. Note that the number of samples in each dataset is annotated in the parenthesis in Fig. 6.

6. CONCLUSION

A no-reference sharpness metric is proposed and verified efficient for realistic data. It is nothing special, but comprises a set of nonlinear statistics on image gradients. The assorted of statistics are closely related to various aspects of pooling strategy; the operators maximum, variance, and kurtosis accentuate the steepest gradients, the pixel-wise average, patch-wise average, and pyramid analysis merge gradients in multi-scales, and the moments based on image composition weigh gradients with visual saliency, while the operators minimum and vector mean attribute gradients to the outcome of different blur type. It is important that those statistics combination is “selected” by a statistic modelling on data, more than a mere handcraft design. Nonetheless, perception on blur involves with high-level vision and goes beyond the proposed statistics. There is still substantial room to improve the measurement by incorporating high-level features.



Figure 7 Failure cases of sharpness overestimated (top) and underestimated (bottom) in our database.

REFERENCES

- [1] Bae, S.& Durand, F. “Defocus magnification”. In *Computer Graphics Forum* (2007), vol. 26,Wiley Online Library, pp. 571–579.
- [2] Bahrami, K. & Kot, A. “A fast approach for no reference image sharpness assessment based on maximum local variation”. *Signal Processing Letters, IEEE* 21, 6 (June 2014), 751–755.
- [3] Brill, M. H., Lubin, J., Costa, P., Wolf, S., & Pearson, J. “Accuracy and cross-calibration of video quality metrics: new methods from atis/t1a1”. *Signal Processing: Image Communication* 19, 2 (2004), 101–107. 6
- [4] Caviedes, J. & Gurbuz, S. “No-reference sharpness metric based on local edge kurtosis”. In *Image Processing Proceedings. International Conference on* (2002), vol. 3, IEEE, pp. III–53.
- [5] Chakpabarti, A., Zickler, T. & Freeman, W. T. “Analyzing spatially-varying blur”. In *Computer Vision and Pattern Recognition (CVPR)*, (2010), IEEE, pp. 2512–2519.
- [6] Chandler, D. & Hemami, S. *A57 database*, 2007.
- [7] Chen, M.-J. & Bovik, A. C. “No-reference image blur assessment using multiscale gradient”. *EURASIP Journal on Image and Video Processing* 2011, 1 (2011), 1–11.

- [8] Chen, X., Yang, J., Wu, Q., Zhao, J. & He, X. "Directional high-pass filter for blurry image analysis". *Signal Processing: Image Communication* 27, 7 (2012), 760–771.
- [9] Chiang, M.-C. & Boulton, T. E. "Local blur estimation and super-resolution". In *IEEE Conference on Computer Vision and Pattern Recognition (1997)*, IEEE Computer Society, pp. 821–821.
- [10] Ciancio, A., Da Costa, A. T., Da Silva, E. A., Said, A., Samadani, R. & Obrador, P. "No-reference blur assessment of digital pictures based on multifeature classifiers". *Image Processing, IEEE Transactions on* 20, 1 (2011), 64–75.
- [11] Dai, S. & Wu, Y. "Removing partial blur in a single image". In *Computer Vision and Pattern Recognition, CVPR. IEEE Conference on (2009)*, IEEE, pp. 2544–2551.
- [12] Gennery, D. B. "Determination of optical transfer function by inspection of frequency-domain plot". *JOSA* 63, 12 (1973), 1571–1577.
- [13] Groen, F. C., Young, I. T. & Lightart, G. "A comparison of different focus functions for use in autofocus algorithms". *Cytometry* 6, 2 (1985), 81–91.
- [14] Hassen, R., Wang, Z. & Salam, M. M. "Image sharpness assessment based on local phase coherence". *Image Processing, IEEE Transactions on* 22, 7 (2013), 2798–2810.
- [15] Jang, S.-I., Chung, J., Lee, Y., Chung, K., Kim, W. & Lee, C.-W. "A real-time identification method on motion and out of focus blur for a video camera". *Consumer Electronics, IEEE Transactions on* 40, 2 (1994), 145–153.
- [16] Ji, H. & Liu, C. "Motion blur identification from image gradients. In *Computer Vision and Pattern Recognition*", *CVPR IEEE Conference on (2008)*, IEEE, pp. 1–8.
- [17] Larson, E. & Chandler, D. *Categorical image quality (CSIQ) database 2009*.
- [18] Le Callet, P., Atrousseau, F., et al. *Subjective quality assessment IRCCyN/IVC database*.
- [19] Liu, R., Li, Z. & Jia, J. "Image partial blur detection and classification". In *Computer Vision and Pattern Recognition, IEEE Conference on (2008)*, IEEE, pp. 1–8.
- [20] Pertuz, S., Puig, D. & Garcia, M. A. "Analysis of focus measure operators for shape from focus". *Pattern Recognition* 46, 5 (2013), 1415–1432.
- [21] Ponomarenko, N., Ieremeiev, O., Lukin, V., et al. "Color image database TID2013: Peculiarities and preliminary results". In *Visual Information Processing (EUVIP), 4th European Workshop on (2013)*, IEEE, pp. 106–111.
- [22] Ponomarenko N., Lukin, V., Zelensky, A., et al. TID2008 - a database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern Radio Electronics* 10, 4 (2009), 30–45.
- [23] Rustin Ol, M., Chazalon, J. & Ogier, J.-M. "Combining focus measure operators to predict ocr accuracy in mobile-captured document images". In *Document Analysis Systems (DAS), 11th IAPR International Workshop on (2014)*, IEEE, pp. 181–185.
- [24] Sheikh, H. R., Wang, Z., Cormack, L. & Bovik, A. C. *Live image quality assessment database release 2, 2005*
- [25] Shi, J., Xu, L. & Jia, J. "Discriminative blur detection features". In *Computer Vision and Pattern Recognition, IEEE Conference on (2014)*, IEEE.
- [26] Su, B., Lu, S. & Tan, C. L. "Blurred image region detection and classification". In *Proceedings of the 19th ACM international conference on Multimedia (2011)*, ACM, pp. 1397–1400.

- [27] Tang, H., Joshi, N. & Kapoor, A. "Blind image quality assessment using semi supervised rectifier networks". In Computer Vision and Pattern Recognition, IEEE Conference on (2014), IEEE.
- [28] Wu, S., Lin, W., Xie, S., Lu, Z., Ong, E. P. & Yao, S. "Blind blur assessment for vision-based applications". Journal of Visual Communication and Image Representation 20, 4 (2009), 231–241.
- [29] Ye, P., Kumar, J., Kang, L. & Doermann, D. "Real-time no-reference image quality assessment based on filter learning". In Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on (2013), IEEE, pp. 987–994.
- [30] Zari, A., Tatalovi, N., Brajkovi, N., et al. FER image quality assessment database. AUTOMATIKA: 53, 4 (2012), 344–354. 6
- [31] Zhang, N. F., Postek, M. T., Larrabee, R. D., et al., "Image sharpness measurement in the scanning electron microscope (part iii)". Scanning 21, 4 (1999), 246–252.
- [32] Zhuo, S. & Sim, T. "Defocus map estimation from a single image". Pattern Recognition 44, 9 (2011), 1852–1858.