Blind image sharpness assessment based on local contrast map statistics☆

Goran Gvozden*, Sonja Grgic, Mislav Grgic

Department of Wireless Communications, Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3/XII, 10000 Zagreb, Croatia

ARTICLE INFO

ABSTRACT

This paper presents a fast blind image sharpness/blurriness assessment model (BISHARP) which operates in spatial and transform domain. The proposed model generates local contrast image maps by computing the root-mean-squared values for each image pixel within a defined size of local neighborhood. The resulting local contrast maps are then transformed into the wavelet domain where the reduction of high frequency content is evaluated in the presence of varying blur strengths. It was found that percentile values computed from sorted, level-shifted, high-frequency wavelet coefficients can serve as reliable image sharpness/blurriness estimators. Furthermore, it was found that higher dynamic range of contrast maps significantly improves model performance. The results of validation performed on seven image databases showed a very high correlation with perceptual scores. Due to low computational requirements the proposed model can be easily utilized in real-world image processing applications.

1. Introduction

A modern information society is overwhelmed with huge amounts of visual content being generated, stored and shared on a daily basis. Each processing phase, from image recording to reproduction, introduces various distortions which can lead to reduced visual experience. One of such distortions, closely related to the loss of visual acuity, limited contrast sensitivity and perceived image sharpness is blur [1]. In certain cases, the blur distortion can question the performance of a human or artificial visual system [2]. On the other side, in certain areas of image processing, the introduction of artificial blur can bring a more realistic virtual environment as well as enhanced visual experience [3]. Methods capable to detect, estimate or classify various types of blur distortion have received increasing attention in the field of image segmentation [4], blur estimation [5,6], sharpness assessment [7] and deblurring [8]. Regardless of the application, an efficient computational model capable to quantify the blur distortion and thus, estimate the image sharpness, can certainly optimize the quality of proliferating visual services.

The human visual system (HVS) is the most reliable estimator of image sharpness; however, the complex and time-consuming evaluation process makes it inapplicable in the real-world image processing environment [9]. The solution to this problem lies in the objective models capable to automatically estimate image sharpness [10]. This objective assessment ecosystem consists of full-reference (FR), reduced reference (RR) and no-reference (NR) methods where differentiation is based on

the availability of reference or undistorted image [11]. Unlike FR and RR methods, no-reference or blind image assessment methods predict the image sharpness without the need for reference image. Accordingly, the NR methods are the most interesting due to their applicability in real-time visual systems where reference images are rarely accessible [12].

The early work in blind image sharpness assessment was mainly performed in the spatial domain. It was based on measurement of edge widths [13] where additional improvements were achieved by using perceptual features based on just noticeable blur (JNB) [14] and cumulative probability of blur detection (CPBD) [15]. Bahrami and Kot [16] estimate sharpness by measuring the spread of maximum local variation (MLV) coefficients. In [17] authors use the local Michelson contrast and energy map elements in an autoregressive (AR) space. Additional gains were demonstrated using the general-purpose image quality assessment (IQA) methods based on neural networks [18,19] and machine learning [20]. In [21] image sharpness is measured using the block energy of sparse coefficients normalized with block variances. Li et al. [22] exploit the Tchebichef moments computed from the image gradient maps, while authors in [23,24] take the Singular Value Decomposition (SVD) approach to assess the image sharpness. The sharpness methods in transform domain observe the statistics and energy information of Discrete Cosine Transform (DCT) coefficients to quantify the strength of blur distortion [25,26]. Some authors observe and parameterize Steerable Pyramid Wavelet Transform (SPWT) coefficient distributions in order to extract the statistics and form a feature

vector relevant for sharpness assessment [27]. Authors in [28–30] observe the local phase coherence of an image to measure the blurriness. In [31], authors proposed a sharpness index expressed as a weighted sum of sub-band log-energies computed in a multi-scale Discrete Wavelet Domain (DWT). Finally, hybrid methods observe and combine the blur features extracted from the spatial and transform domain [32]. Li et al. [33] employ sharpness features extracted from multi-scale spatial/frequency domains in combination with the machine learning model, while Zhang et al. [34] use NSS based features with learned multivariate Gaussian model to measure other image distortions as well.

Our goal is to build an efficient blind computational model capable to: identify the blur distortion signature, quantify the level of distortion, and estimate the perceived image sharpness/blurriness. We observe the influence of isotropic blur on local contrast statistics in the spatial and frequency domain as well as its implications on perceived image sharpness/blurriness. Contrast has been used as a constituent feature in many IQA models. As part of the structural similarity paradigm, the local standard deviation based contrast was used within full reference models operating in single-scale [35] and multi-scale spatial domain [36]. In the blind image assessment environment, changes in contrast were accounted by measuring the total [32] or maximum local variation in image intensities [16]. The gradient magnitudes (GM) and Laplacian of Gaussian (LOG) responses were also successfully employed as local spatial contrast features in [37]. Some authors use locally computed Michaelson contrast in the autoregressive space [17], while others observe the transform domain to extract DCT [38] or DWT [39] based local contrast features.

Our model builds upon the perceptual contrast sensitivity mechanism which has a fundamental role in the visual information processing, especially in terms of discrimination of spatial and temporal patterns falling onto the retina. We use the local root-mean-square

(RMS) contrast measure, to generate contrast image maps and capture the intensity variations across the image [40,41]. Unlike previous approaches, we introduce the concept of increased dynamic range. By increasing the dynamic range of generated contrast maps the performance of proposed model is significantly improved. Furthermore, we extended our approach into the high frequency discrete wavelet space due to the well-known fact that high frequency image components are attenuated by blur [42]. Here, the level-shifting operation, performed on decomposed wavelet coefficients, is introduced to further enhance the model performance. The computed statistical parameters – in our case percentile values – are defined as perceptually significant image sharpness/blurriness features. The proposed model demonstrates very high correlation with perceptual scores. If processing time in combination with prediction accuracy across all tested databases is taken into consideration, the proposed method outperforms other state-of-the-art image sharpness metrics. These findings were validated as part of an extensive performance evaluation, which involved seven publicly available blur databases.

The following section describes the proposed algorithm. In Section 3 the results of performance evaluation are given along with short description of databases and performance attributes used for model validation. Additionally, a thorough evaluation of the proposed sharpness measure is presented including results of correlation analysis, hypothesis testing and computational complexity. In Section 4, the paper ends with the concluding remarks and future work.

2. Image sharpness assessment model

HVS encodes only relative luminance values based on its intrinsic light adaptation mechanism [43]. The processing of visual information is performed using a nonlinear function where only values that show 1% change in luminance are being registered and coded. This capability expressed as the contrast sensitivity represents one of the fundamental attributes of visual perception. Moreover, the perceptual response to an image strongly depends on image contrast ratio which is commonly defined as the ratio between the maximum and minimum luma[1] values [44]. In the context of image quality assessment the contrast is found to be inextricably related to image sharpness where images with higher contrast ratios are perceived as sharper [39]. Hence, observing and quantifying the changes in contrast can contribute to reliable sharpness/blurriness estimation [36,45]. Next, we explain how local contrast is computed, integrated and utilized within our sharpness model.

2.1. Blind image sharpness estimation framework

The proposed BISHARP model incorporates processing in the spatial and wavelet transform domain. The flowchart of the proposed model is shown in Fig. 1. It is a fast and straightforward process where an image being tested for sharpness is first converted to grayscale domain. Then, the local contrast map is generated computing the root mean square values in local pixel neighborhood. The generated map is transformed to frequency domain using one-scale discrete wavelet transform. Extracted sub-band coefficients are sorted and level-shifted by the maximum value found in a negative valued wavelet coefficients pool. The computed percentile value of the resulting, level-shifted wavelet coefficients distribution represents the final image sharpness score. Below, we present the 5-step framework designed to compute the image sharpness measure.

2.1.1. Conversion to grayscale image

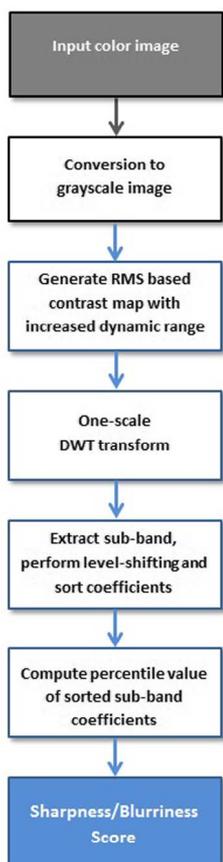**Step 1**. The first step is conversion from color *R′G′B′* to grayscale

Fig. 1. Flowchart of the proposed BISHARP model.

---

[1] Luma is defined as a weighted sum of tristimulus R′G′B′ values obtained after processing the linear RGB values with nonlinear gamma function. Gamma function is an approximation of the perceptual response to luminance.
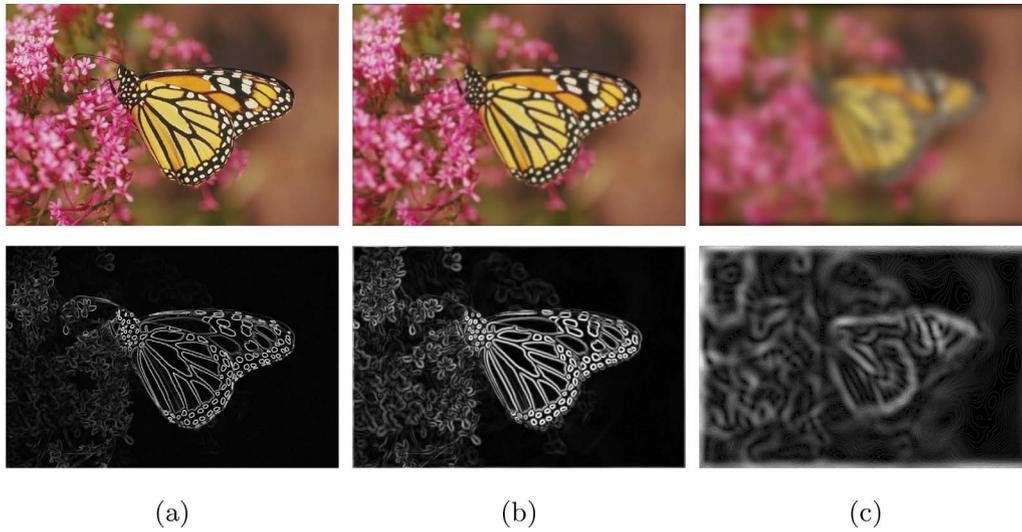
Fig. 2. The first row depicts undistorted image Monarch (a) and its blurred versions, (b) and (c). The second row depicts corresponding LCM_{rms} local contrast maps.

(a)   (b)   (c)

images. The conversion is performed according to the expression standardized in Rec. BT-601 [46] and defined in Matlab [47] as

$$I = 0.299R' + 0.587G' + 0.114B'. \tag{1}$$

2.1.2. Computation of local contrast

Step 2. It is well known that image regions populated with high intensity variations carry more information weight, which in turn attract more visual attention [48]. Here, we employ the local RMS contrast measure to capture these variations and measure the effects of varying blur levels on image sharpness. Moreover, the additional reason to use the RMS contrast lies in its ability to simulate the perceptual sensitivity to contrast – in particular the perceived contrast in natural images [49,50]. Local RMS values were brought to focus to address the perceptual contrast response and contrast gain control mechanisms that occur during saccadic inspection of natural images [41]. Hence, the local contrast map is computed as

$$LCM_{rms}(x,y) = \sigma_{loc}(x,y)^{\alpha}, \tag{2}$$

where $\sigma_{loc}(x,y)$ is the local standard deviation of image intensities found in a neighborhood centered around a pixel at location (x,y). The dynamic range of the map is increased by raising each element of the standard deviation map $\sigma_{loc}(x,y)$ to the power of α. The standard deviation in a local neighborhood is computed as

$$\sigma_{loc}(x,y) = \sqrt{\frac{1}{mn} \sum_{k=-K}^{K} \sum_{l=-L}^{L} (I(x+k,y+l) - \mu_{loc}(x,y))^2}, \tag{3}$$

where $I(x,y)$ is the intensity value of image pixel at location (x,y), $m = 2|K| + 1$ is the size of local window in vertical direction, $n = 2|L| + 1$ is the size of the local window in horizontal direction and the total number of pixels within a window is $m \times n$. The local window mean value is computed as

$$\mu_{loc}(x,y) = \frac{1}{mn} \sum_{k=-K}^{K} \sum_{l=-L}^{L} (I(x+k,y+l)). \tag{4}$$

Border pixels were computed using symmetric padding. Fig. 2 depicts the original and distorted versions of Monarch.bmp image along with their corresponding RMS local contrast image arrays. The elements of the new LCM_{rms} image array, as shown in the second row of Fig. 2, were sorted in a descending order to observe the behavior in the presence of blur. Fig. 3 depicts distributions of LCM_{rms} generated map coefficients for image Monarch.bmp distorted with different strengths of Gaussian blur ranging from $\sigma = 0.9062$ to $\sigma = 11.3333$. Images are rated with MOS[2] scores ranging from 23.2355 ($\sigma = 0.9062$) for the least distorted image to 75.9245 ($\sigma = 11.3333$) for the image with the highest

blur distortion. The comparison in the spatial domain clearly differentiates the distributions based on the strength of applied blur.

2.1.3. Transformation to wavelet space

Step 3. Edges are important image features characterized by high intensity discontinuities and variations. These variations represent the high energy content being attenuated in the presence of blur. The same applies to LCM_{rms} image array where the intensity discontinuities and the changes in frequency content are even more pronounced. In order to capture this reduction in energy at high frequencies we moved from spatial to frequency domain by performing the two-dimensional multi-scale DWT transform. Hence, with this computationally fast image transformation we captured, observed and analyzed the energy reduction at different scales and orientations in order to extract an image feature sensitive to blur distortion. Decomposition of contrast maps was performed using symmetric quadrature mirror filters with 9-tap filter kernels, qmf9 = {0.02807382, −0.060944743, −0.073386624, 0.41472545, 0.7973934, 0.41472545, −0.073386624, −0.060944743, 0.02807382}, as defined in [51].

Hence, the local contrast maps are convolved with odd, symmetric filter where boundary handling was performed by reflecting the boundary pixels. Given LCM_{rms} map is decomposed into low frequency band made of approximation coefficients and three high frequency sub-bands containing horizontal, vertical and diagonal wavelet coefficients, respectively. Resulting horizontal, vertical and diagonal high-frequency sub-bands are denoted as W_{hn}, W_{vn}, W_{dn} at decomposition level n. Our model performs one-scale decomposition where the high frequency diagonal coefficients are taken and used in the following, fourth step of sharpness feature extraction process. Fig. 4 shows how diagonal high frequency sub-band coefficients are being attenuated in the presence of varying blur strengths.

2.1.4. Transforming level-shifted and sorted coefficients to log domain

Step 4. Next, we follow the concept proposed in [29] which puts more emphasis on sharper image regions by ranking and weighting the wavelet coefficients. However, before sorting the coefficients we perform level-shifting to accentuate the difference between different mean blur levels. The level-shifting is performed by finding the minimum value in the diagonal wavelet coefficient set, min{W_{d1}}; taking its absolute value, abs(min{W_{d1}}) and adding it to all wavelet coefficient magnitudes W_{d1_i} according to
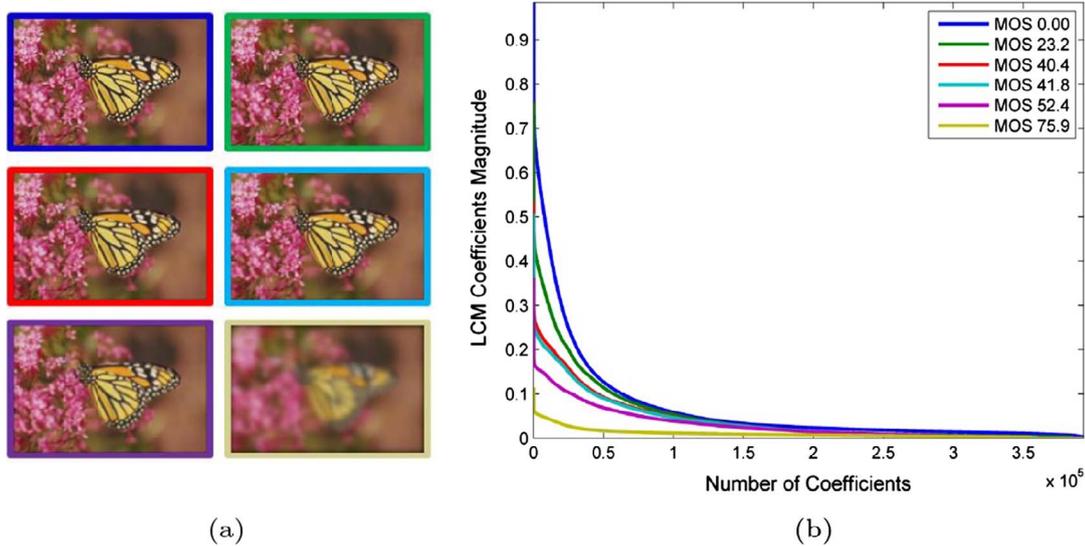

Fig. 3. Comparison of sorted LCM_{rms} coefficient distributions obtained using 3-by-3 window size. (a) Original Monarch image and its distorted versions. (b) LCM_{rms} coefficient distributions corresponding to images on the left.
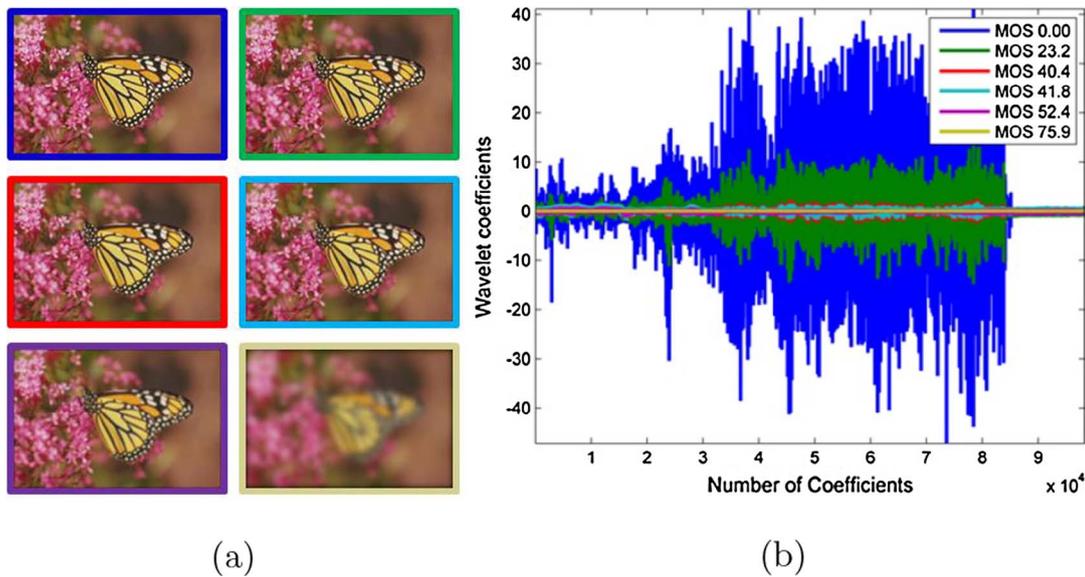

Fig. 4. (a) Original Monarch.bmp image and its distorted versions framed with different colors corresponding to different levels of blur distortion. (b) Distributions of high frequency diagonal wavelet coefficients (W_{d1}) for original and distorted versions of Monarch.bmp. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$W_{D1_i} = W_{d1_i} + abs(min(W_{d1})) \quad i = 1,...,N. \tag{5}$$

After sorting, the resulting level-shifted and ranked coefficient distribution is described as $\{W_{D1_i} \mid i = 1, 2, ..., N\}$, where $W_{D1_1} \geqslant W_{D1_2} \geqslant \cdots \geqslant W_{D1_N}$ and N represents the total number of wavelet coefficients found in the diagonal sub-band. Since increased dynamic range implies high wavelet coefficient values, the log transformation using log_{10} function was applied to each coefficient. Thus, the existing ranked collection of wavelet coefficients was transformed to log domain producing the new collection of coefficients $\{log_{10} W_{D1_i} \mid i = 1, 2, ..., N\}$, where $log_{10} W_{D1_1} \geqslant log_{10} W_{D1_2} \geqslant \cdots \geqslant log_{10} W_{D1_N}$. In Fig. 5(b) we show the result of level-shifting performed on coefficient distributions shown in Fig. 5(a). Fig. 5(c) depicts sorted wavelet coefficient distributions and Fig. 5(d) shows the sorted coefficients in log domain along with the extraction of sharpness measure.

2.1.5. Computation of percentile value

Step 5. At the final stage we compute the percentile value of ranked

distribution that corresponds to the blind image sharpness score. A percentile is the coefficient value below which a certain percentage (%) of level-shifted and ranked diagonal wavelet coefficients lie. It was found that the highest susceptibility to blur distortion is in lower percentile range. In our case, the 0.25th percentile was used to compute the BISHARP score. The sharpness score is computed according to

$$BISHARP = log_{10} W_{D1}(i), \tag{6}$$

where i is the rank of wavelet coefficient that is higher than 0.25% of the sorted wavelet coefficients. The Matlab's *prctile* function was applied to compute the objective scores.

Algorithm 1. The proposed BISHARP algorithm

> **Data:** Input color image I
> **Result:** Sharpness score BISHARP
> **Initialization:** Set $\alpha = 3.75$, p = .25, window size n = 3 pixels
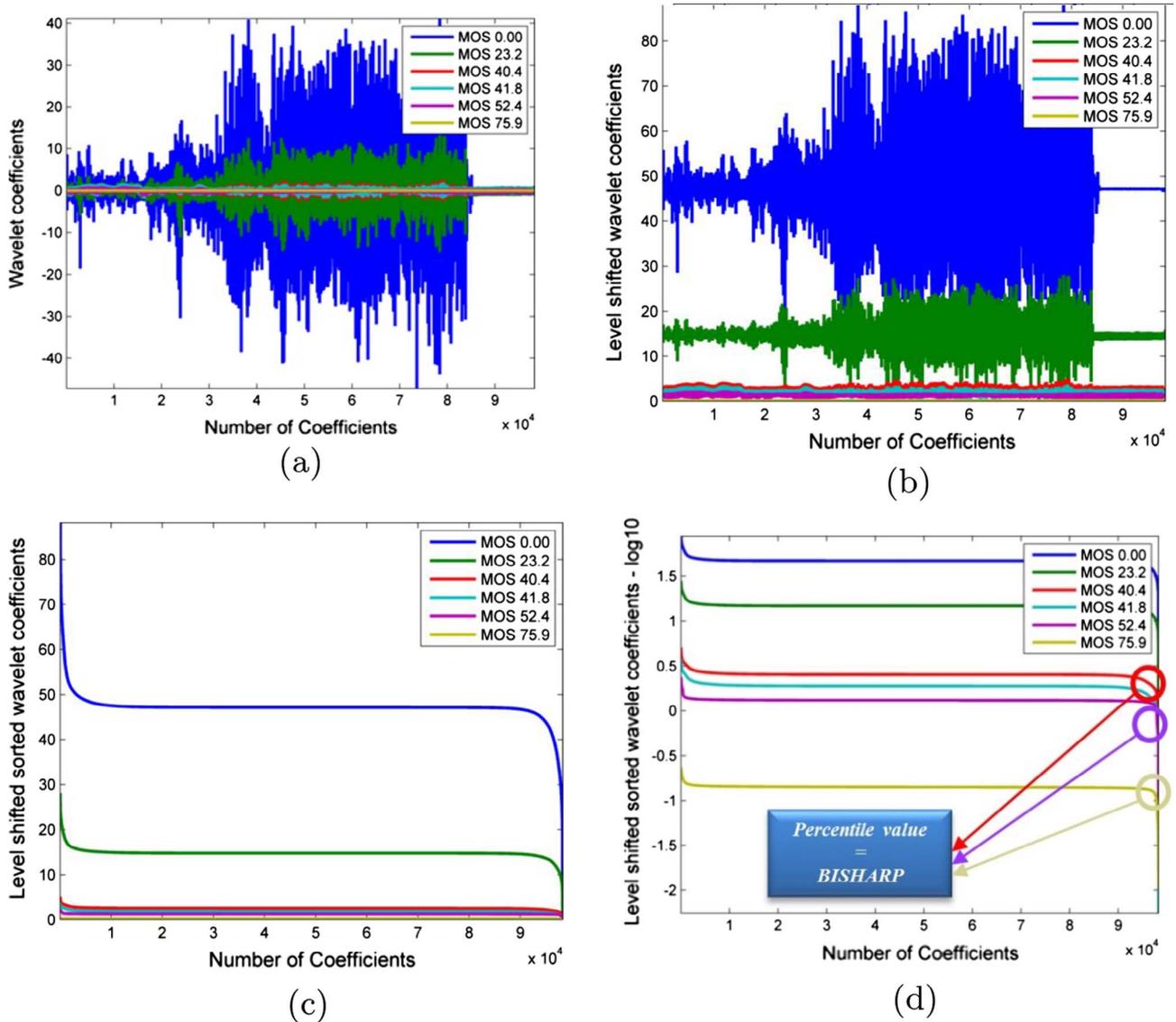> **Step 1:** Convert RGB to grayscale image using Eq. (1)

Fig. 5. (a) High frequency wavelet coefficient distributions corresponding to varying blur strengths. (b) Level-shifted high frequency diagonal coefficient distributions (W_{D1}). (c) Sorted level-shifted coefficient distributions. (d) Coefficient distributions in the log domain.

Step 2: Compute the increased dynamic range LCM_{rms} map using Eq. (2)

Step 3: Compute DWT transform to extract diagonal wavelet coefficients

Step 4: Transform the level-shifted (Eq. (5)) and sorted diagonal wavelet coefficients to log10 domain

Step 5: Compute sharpness score BISHARP using Eq. (6)

In this section we observed how blur affects the distribution of local contrast map elements in both, spatial and frequency domain. Based on this analysis, we have extracted the sharpness features susceptible to varying levels of blur distortion. The percentile value of ranked, level-shifted, high frequency wavelet coefficients is defined as a perceptually significant feature capable to estimate image sharpness. The sharpness score based on local RMS contrast map is summarized in Algorithm 1.

3. Experimental results and discussion

In this section we give an overview of the testing environment and performance metrics necessary to validate the proposed model. Finally, the results of hypothesis testing are presented to show the relationship among tested methods.

3.1. Image databases

The proposed model was evaluated on seven publicly available databases. The first blurred image dataset consisted of images found in six conventional databases including LIVE Image Quality Assessment Database (Release 2) [52,53], Categorical Subjective Image Quality Database – CSIQ [54], Tampere Image Database 2008 – TID2008 [55], Tampere Image Database 2013 – TID2013 [56], VCL@FER image quality assessment database [57] and IRCCyN/IVC Image Quality Database [58]. It is worth noting that blurred images within LIVE, TID2008 and TID 2013 database were introduced with the wraparound error due to Gaussian filtering operation applied to reference images. This wraparound error was removed by cropping the image edges by 6 pixels. The second blurred image dataset was taken from Waterloo Exploration Database [59] which consists of 4744 original images and 28464 images blurred at five evenly distributed distortion levels. Unlike other databases, the subjective scores of images were not provided; however, three new test criteria were proposed to test the model robustness and generalization capability in the real-world environment.

Table 1
Results of performance evaluation on six conventional databases.

Database	LIVE					CSIQ						
Parameter	PLCC	SROCC	KROCC	RMSE	MAE	PLCC	SROCC	KROCC	RMSE	MAE	OR	OD
NIQE	*0.9246*	*0.9427*	*0.8011*	*8.2864*	*6.2570*	*0.8832*	*0.8736*	*0.6981*	*0.1344*	*0.1007*	*0.3267*	*3.9053*
ILNIQE	*0.8850*	*0.9109*	*0.7492*	*10.1279*	*7.9976*	*0.7039*	*0.8295*	*0.6430*	*0.2035*	*0.1529*	*0.4400*	*9.8805*
SSEQ	*0.9574*	*0.9503*	*0.8141*	*6.2823*	*5.0238*	*0.8327*	*0.8395*	*0.6494*	*0.1587*	*0.1262*	*0.3533*	*5.8482*
dipIQ	*0.9356*	*0.9512*	*0.8112*	*7.6775*	*5.9420*	*0.9203*	*0.8997*	*0.7294*	*0.1121*	*0.0805*	*0.2267*	*2.4871*
SPARISH	*0.9680*	*0.9727*	*0.8626*	*5.4585*	*4.1214*	*0.8603*	*0.8864*	*0.7131*	*0.1734*	*0.1396*	*0.4133*	*6.4552*
LPC-SI	0.9178	0.9223	0.7668	8.6362	6.7060	0.9213	0.9024	0.7362	0.1115	0.0821	0.26667	2.4304
MLV	0.9630	0.9575	0.8293	5.8620	4.6623	0.8759	0.9074	0.7435	0.1383	0.1081	0.3600	5.1041
CPBD	0.9041	0.9360	0.7915	9.2953	7.0974	0.8102	0.8636	0.6817	0.1680	0.1346	0.3867	6.9381
S3	0.9462	0.9438	0.8015	7.0353	5.6260	0.7703	0.8860	0.7017	0.1827	0.1495	0.4467	8.8776
FISH	0.9191	0.9278	0.7674	8.5696	6.8382	0.8516	0.8703	0.6997	0.1502	0.1179	0.3667	5.5567
FISH$_{bb}$	0.9557	0.9603	0.8353	6.3998	5.1050	0.8929	0.8939	0.7262	0.1290	0.1004	0.3133	3.7185
ARISM	0.9616	0.9666	0.8419	5.9673	4.7008	0.9053	0.8999	0.7265	0.1217	0.0879	0.2733	2.9064
ARISM$_c$	0.9648	0.9699	0.8509	5.7235	4.4634	0.9046	0.9056	0.7362	0.1221	0.0890	0.2533	2.9625
BISHARP	0.9614	0.9611	0.8388	5.9818	4.4647	0.9186	0.9125	0.7485	0.1132	0.0851	0.2733	2.6321

Database	VCL@FER					TID2008						
	PLCC	SROCC	KROCC	RMSE	MAE	PLCC	SROCC	KROCC	RMSE	MAE	OR	OD
NIQE	*0.8870*	*0.8913*	*0.7038*	*11.2468*	*8.8462*	*0.7844*	*0.7711*	*0.5586*	*0.7279*	*0.5732*	*0.7600*	*36.1733*
ILNIQE	*0.8338*	*0.8491*	*0.6496*	*13.4456*	*10.7660*	*0.8341*	*0.8258*	*0.6095*	*0.6473*	*0.5105*	*0.7100*	*30.7650*
SSEQ	*0.8977*	*0.8839*	*0.7013*	*10.7298*	*8.2127*	*0.8109*	*0.8030*	*0.6051*	*0.6867*	*0.5320*	*0.7200*	*32.7538*
dipIQ	*0.9416*	*0.9402*	*0.7785*	*8.2022*	*6.7147*	*0.8940*	*0.8971*	*0.7199*	*0.5258*	*0.4178*	*0.6700*	*22.1713*
SPARISH	*0.9388*	*0.9302*	*0.7622*	*8.3879*	*0.6556*	*0.8877*	*0.8844*	*0.6969*	*0.5403*	*0.4109*	*0.6600*	*21.9538*
LPC-SI	0.9161	0.9147	0.7470	9.7662	7.3332	0.8921	0.8960	0.7155	0.5302	0.4188	0.6300	21.6874
MLV	0.8901	0.8791	0.7097	11.1034	8.1238	0.8558	0.8528	0.6508	0.6071	0.4669	0.6800	26.7230
CPBD	0.9289	0.9227	0.7482	9.0217	7.2636	0.7644	0.7774	0.5691	0.7566	0.5763	0.6800	37.3372
S3	0.9048	0.8507	0.6619	10.3739	8.3317	0.7994	0.7709	0.5351	0.7051	0.5431	0.6800	33.0813
FISH	0.8865	0.8784	0.6936	11.2688	8.4029	0.8074	0.7922	0.5630	0.6923	0.5517	0.7000	34.6953
FISH$_{bb}$	0.9218	0.9120	0.7381	9.4448	7.0956	0.8378	0.8255	0.6014	0.6407	0.5044	0.6800	29.8311
ARISM	0.9432	0.9259	0.7561	8.0918	6.7420	0.8358	0.8426	0.6443	0.6444	0.5019	0.7100	29.4547
ARISM$_c$	0.9462	0.9306	0.7626	7.8852	6.5196	0.8502	0.8622	0.6665	0.6177	0.4858	0.7300	27.7392
BISHARP	0.9066	0.8976	0.7180	10.2797	7.8229	0.8911	0.8850	0.6932	0.5326	0.4157	0.6300	22.0558

Database	IVC					TID2013						
	PLCC	SROCC	KROCC	RMSE	MAE	PLCC	SROCC	KROCC	RMSE	MAE	OR	OD
NIQE	*0.8078*	*0.8036*	*0.6455*	*0.6729*	*0.5298*	*0.8230*	*0.7692*	*0.5649*	*0.7089*	*0.5747*	*0.7680*	*45.6378*
ILNIQE	*0.9074*	*0.8510*	*0.6773*	*0.4797*	*0.4065*	*0.8396*	*0.8318*	*0.6258*	*0.6779*	*0.5489*	*0.7440*	*42.9001*
SSEQ	*0.9026*	*0.8239*	*0.6667*	*0.4914*	*0.3793*	*0.8251*	*0.8219*	*0.6227*	*0.7050*	*0.5588*	*0.7440*	*44.0241*
dipIQ	*0.8112*	*0.7585*	*0.5185*	*0.6676*	*0.5195*	*0.9201*	*0.9178*	*0.7464*	*0.4886*	*0.3846*	*0.6080*	*24.7042*
SPARISH	*0.9403*	*0.9398*	*0.8042*	*0.3885*	*0.3180*	*0.9006*	*0.8916*	*0.7014*	*0.5424*	*0.4323*	*0.6400*	*29.3439*
LPC-SI	0.9700	0.9564	0.8466	0.2776	0.1931	0.9165	0.9202	0.7484	0.4992	0.3943	0.5920	25.7714
MLV	0.9777	0.9767	0.8889	0.2399	0.1743	0.8802	0.8762	0.6795	0.5921	0.4640	0.6720	32.6495
CPBD	0.8011	0.7690	0.6138	0.6834	0.5151	0.8195	0.8170	0.6067	0.7151	0.5646	0.7280	44.8457
S3	0.8271	0.8691	0.7090	0.6417	0.5245	0.8482	0.8269	0.5992	0.6609	0.5248	0.7040	40.1041
FISH	0.9585	0.9323	0.7937	0.3256	0.2658	0.8327	0.8102	0.5878	0.6910	0.5735	0.8000	45.1325
FISH$_{bb}$	0.9447	0.9187	0.7513	0.3742	0.2865	0.8662	0.8506	0.6299	0.6236	0.5121	0.7680	37.9583
ARISM	0.8598	0.7863	0.6138	0.5830	0.4274	0.8922	0.8948	0.7094	0.5636	0.4464	0.7120	30.6371
ARISM$_c$	0.8766	0.8089	0.6349	0.5494	0.4032	0.8963	0.8989	0.7136	0.5533	0.4463	0.6960	30.6333
BISHARP	0.9776	0.9797	0.8995	0.2405	0.1771	0.9089	0.9088	0.7254	0.5203	0.4161	0.6400	28.1512

3.2. Evaluation metrics

The performance evaluation of the proposed objective model was conducted according to the Video Quality Experts Group (VQEG) recommendation [60]. A number of performance attributes and corresponding metrics were measured to determine how well the proposed objective method predicts the image sharpness. Based on subjective and objective data sets the prediction monotonicity is determined using Spearman Rank Order Correlation Coefficient (SROCC) and Kendall Rank Order Correlation Coefficient (KROCC). SROCC is defined as

$$SROCC = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N(N^2-1)},$$
(7)

where d_i represents the difference between ranks of subjective and objective scores for *i*-th image, *N* is the number of scores. KROCC is defined according to the following expression:

$$KROCC = \frac{2(N_c - N_d)}{N(N-1)},$$
(8)

where N_c and N_d represent concordant and discordant sample pairs, respectively. Determining other performance attributes such as prediction accuracy and prediction consistency required additional data set adjustment. Hence, before computing prediction accuracy via Pearson Linear Correlation Coefficient (PLCC); prediction consistency via Outlier Ratio (OR); Outlier Distance (OD) as well as other metrics expressed through Root Mean Square Error (RMSE), Mean Absolute Error (MAE); the data fitting using a non-linear regression was performed. The result of the non-linear data fitting is an analytical expression which describes the functional relationship between observed and predicted variables. Here, the 5-parameter logistic function was used

$$y_p(x) = \beta_1\left(\frac{1}{2} - \frac{1}{e^{\beta_2(x-\beta_3)}}\right) + \beta_4 x + \beta_5,$$
(9)

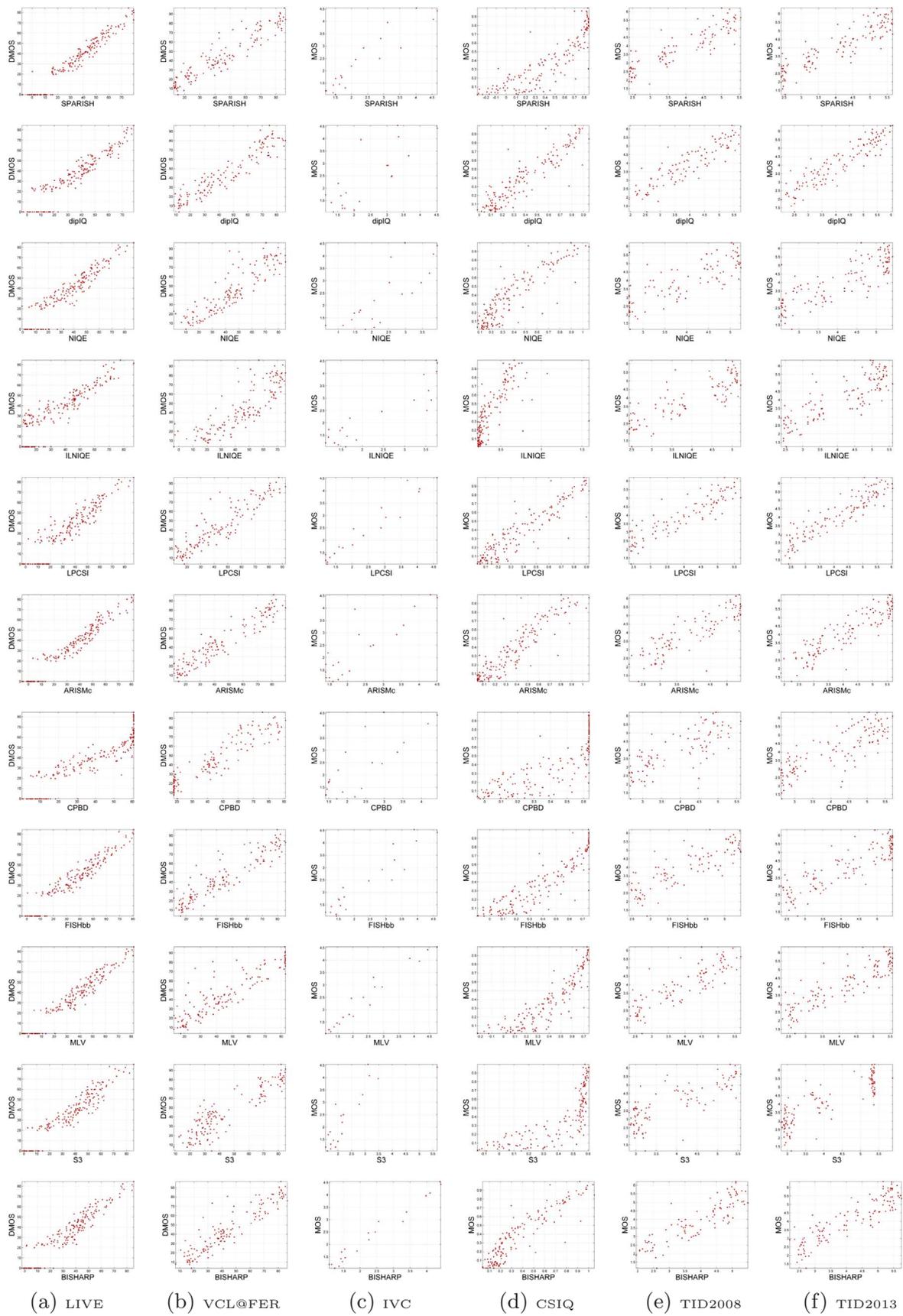(a) LIVE (b) VCL@FER (c) IVC (d) CSIQ (e) TID2008 (f) TID2013

Fig. 6. Scatter plots of subjective versus objective scores.

Table 2
The weighted and direct average results of performance evaluation across all databases.
Two best performing methods were highlighted.

Method	$PLCC_w$	$SROCC_w$	$KROCC_w$	$PLCC_d$	$SROCC_d$	$KROCC_d$
MLV	0.9009	0.9029	0.7377	0.9071	0.9029	0.7503
CPBD	0.8514	0.8699	0.6906	0.8380	0.8476	0.6685
S3	0.8594	0.8661	0.6770	0.8493	0.8579	0.6681
NIQE	0.8674	0.8591	0.6798	0.8517	0.8419	0.6620
IL-NIQE	0.8220	0.8539	0.6636	0.8340	0.8497	0.6591
SSEQ	0.8736	0.8667	0.6896	0.8711	0.8537	0.6765
FISH	0.8685	0.8661	0.6787	0.8760	0.8685	0.6842
$FISH_{bb}$	0.9030	0.8971	0.7214	0.9032	0.8935	0.7137
ARISM	0.9131	0.9092	0.7428	0.8997	0.8860	0.7153
$ARISM_c$	0.9176	0.9162	0.7529	0.9064	0.8960	0.7275
LPC-SI	0.9158	0.9135	0.7482	0.9223	**0.9187**	0.7601
SPARISH	0.9154	0.9183	**0.7576**	0.9160	0.9175	**0.7670**
dipIQ	**0.9214**	**0.9191**	0.7548	0.9038	0.8941	0.7173
BISHARP	**0.9229**	**0.9189**	**0.7571**	**0.9274**	**0.9241**	**0.7706**

where x denotes the objective scores before regression whereas y_p denotes the predicted MOS/DMOS scores after regression; $\beta_1, \beta_2, \beta_3, \beta_4$ and β_5 are parameters depending on y (MOS/DMOS) value range and initial value set depending on the used database. The data set values after non-linear regression are usually depicted on a scatter plot providing a visual representation of the relationship. The PLCC metric measures the prediction accuracy attribute according to the following expression

$$PLCC = \frac{\sum_{i=1}^{N} (y_{pi} - \overline{y_p})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{N} (y_{pi} - \overline{y_p})^2 \sum_{i=1}^{N} (y_i - \overline{y})^2}}, \tag{10}$$

where y_{pi} represents the predicted score for i-th image, $\overline{y_p}$ represents the average of predicted objective scores, y_i is the subjective score for i-th image and $\overline{y}$ is the mean value of all subjective scores. Outlier Ratio measures the prediction consistency according to the following expression

$$OR = \frac{N_o}{N}, \tag{11}$$

where N_o is the number of outliers, data points or y_{pi} values that lie outside of the 95% confidence interval constrained by two standard deviations σ_{y_i}; where σ_{y_i} is a standard deviation of all subjective ratings for a given i-th image. Outliers are defined as values outside of the confidence interval

$$|y_{pi} - \overline{y_i}| > 2\sigma_{y_i}. \tag{12}$$

Furthermore, the outlier distance measure proposed in [54] was used as one of the performance metrics. It measures the distance of an outlier data points from the closest error bar $\pm 2\sigma_{y_i}$. The outlier distance is defined as

$$OD = \sum_{x \in X_o} min \left(|y_{pi} - y_i - 2\sigma_{y_i}|, |y_{pi} - y_i + 2\sigma_{y_i}| \right). \tag{13}$$

where X_o is the set of predicted scores lying outside of confidence interval, y_{pi} is predicted score or outlier for i-th image and y_i is subjective score for i-th image. OD cannot be used for comparison among databases which use different subjective scores metric, e.g. DMOS vs MOS database. RMSE metric measures the average error between subjective and predicted scores providing the information about the variance of data sets

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_{pi} - y_i)^2}, \tag{14}$$

whereas MAE metric is defined as

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_{pi} - y_i|. \tag{15}$$

Additionally, three test criteria introduced in [59] were employed to compare the methods. The first, discriminability test (D-test) evaluates the model's ability to distinguish the original from distorted images. Before determining the model discriminability, indices i of all images in a given database are grouped into sets S_o and S_d representing original and distorted image indices, respectively. Then, threshold T is used to separate images according to $S_o' = \{i | y_{p_i} > T\}$ and $S_d' = \{i | y_{p_i} \leqslant T\}$. Based on this classification the average classification rate R is defined as

$$R = \frac{1}{2} + \left(\frac{|S_o \cap S_o'|}{|S_o|} + \frac{|S_d \cap S_d'|}{|S_d|} \right). \tag{16}$$

With optimized value T and maximum R the discriminability index is defined as

$$D = \max_T R(T). \tag{17}$$

D lies in the interval [0–1] where higher values correspond to better discriminability.

The second test is the listwise ranking consistency test (L-test). This test evaluates the model's ability to monotonically estimate the image sharpness with the increasing strengths of blur distortion. The L-test value represents the mean of all SROCC/KROCC values computed between the total number of distortion levels and the corresponding quality scores for each image in a given database. The L-test using SROCC values is defined as

$$L_S = \frac{1}{SK} \sum_{i=1}^{S} \sum_{j=1}^{K} SROCC(l_{ij}, y_{p_{ij}}), \tag{18}$$

and the L-test using KROCC values is defined as

$$L_K = \frac{1}{SK} \sum_{i=1}^{S} \sum_{j=1}^{K} KROCC(l_{ij}, y_{p_{ij}}), \tag{19}$$

where S is the number of images in a given database, K is the number of distortion types, $y_{p_{ij}}$ represents the set of predicted scores at distortion levels l_{ij} that are obtained by a given model on the same i-th original image and the same j-th distortion type. The third test is the pairwise preference consistency test (P-test). It measures the model's ability to accurately determine the pairwise prediction preference. P-test is defined according to the following expression

$$P = \frac{M_c}{M}, \tag{20}$$

where M is the total number of generated quality-discriminable image pairs (DIPs) in a given database whose absolute MOS/DMOS difference are greater than a predefined threshold T. M_c is the number of concordant pairs found in a DIP set. P lies in the interval [0–1] where higher values correspond to better P-test performance.
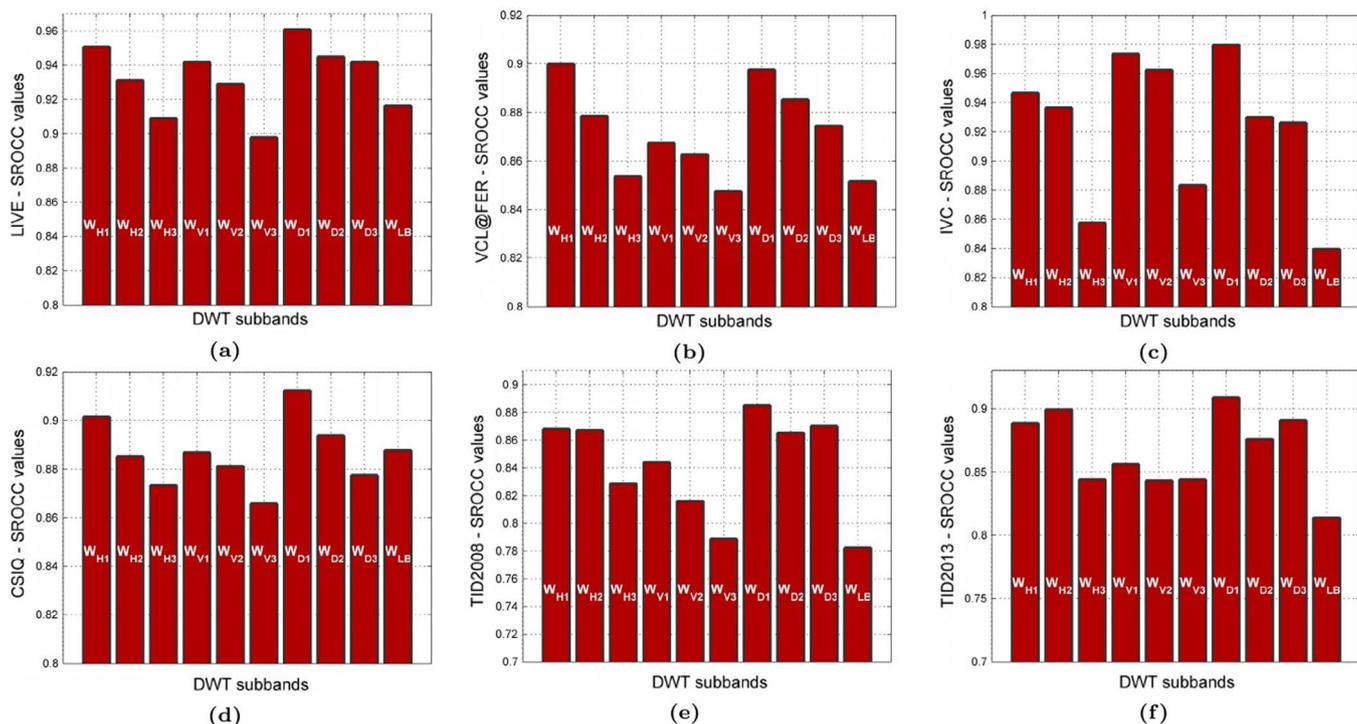
Table 3
Results of average run times obtained while processing the blurred image database.

Method	FISH	BISHARP	MLV	*NIQE*	LPC-SI	CPBD	FISHbb	*SSEQ*	*dipIQ*	*IL-NIQE*	*SPARISH*	ARISMc	S3
Average time (sec)	0.11	0.13	0.17	0.27	1.04	1.14	3.98	5.85	6.32	9.83	15.32	66.76	79.60

Fig. 7. SROCC values obtained at different wavelet sub-bands for (a) LIVE, (b) VCL@FER, (c) IVC, (d) CSIQ, (e) TID2008 and (f) TID2013 database. Sub-bands are ordered from left to right: horizontal – W_{H1}, W_{H2}, W_{H3}; vertical – W_{V1}, W_{V2}, W_{V3}; diagonal – W_{D1}, W_{D2}, W_{D3}; low frequency sub-band at third level (W_{A3}).
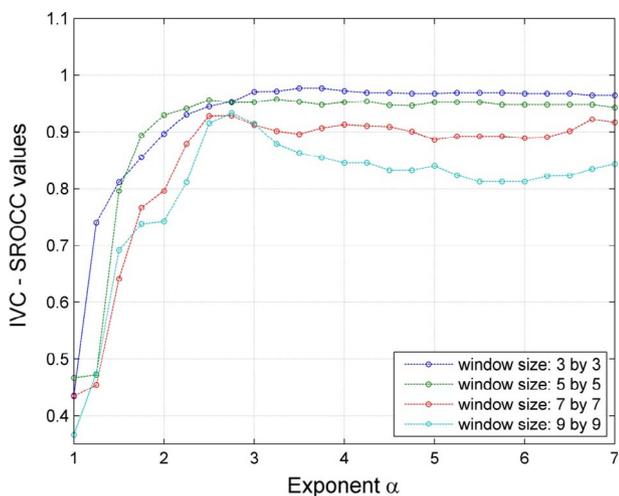

Fig. 8. SROCC values as a function of varying local window size and exponent α for diagonal sub-band extracted after DWT transform of RMS based local contrast maps.

Fig. 9. SROCC as a function of different percentile values for TID2013 database.

3.3. Metrics comparison

The proposed method was compared to thirteen blind image sharpness metrics. Our BISHARP[3] method was tested against (1) MLV [16]; (2) CPBD [61]; (3) S3 [32]; (4) LPC-SI [29,30]; (5) SSEQ [62]; (6) FISH [31]; (7) FISH$_{bb}$ – block based FISH [31]; (8) ARISM [17], (9) ARISM$_c$ – ARISM in YIQ color space [17]; (10) SPARISH [21]; (11) dipIQ [19]; (12) NIQE [63] and (13) IL-NIQE [34]. The dipIQ, NIQE, IL-NIQE and SSEQ are learning based methods that belong to the general-purpose image quality metrics designed for other distortions as well. All above mentioned methods and corresponding source codes were publicly available and easily accessible. No modifications were made to the

original Matlab codes provided by authors. The results of the performance evaluation are shown in Table 1. Two best evaluation results are highlighted with boldface, while machine learning methods are italicized. It is worth mentioning that the SPARISH method is also italicized because it uses an overcomplete dictionary trained on the LIVE database. In general, the blind image sharpness metrics based on machine learning tools are sensitive to the image databases used for training [33].

BISHARP demonstrated high prediction accuracy, monotonicity and consistency on all tested databases. In comparison with three general-purpose metrics, namely SSEQ, NIQE and IL-NIQE, the BISHARP model showed to be superior on all six databases. However, the fourth tested general-purpose method, namely dipIQ showed to be slightly better on TID2008, TID2013 and VCL@FER. The best results were obtained on CSIQ database for SROCC/KROCC performance attributes. Performance evaluation on the LIVE database showed slightly worse results in comparison to SPARISH and ARISM$_c$. In case of VCL@FER database the

---

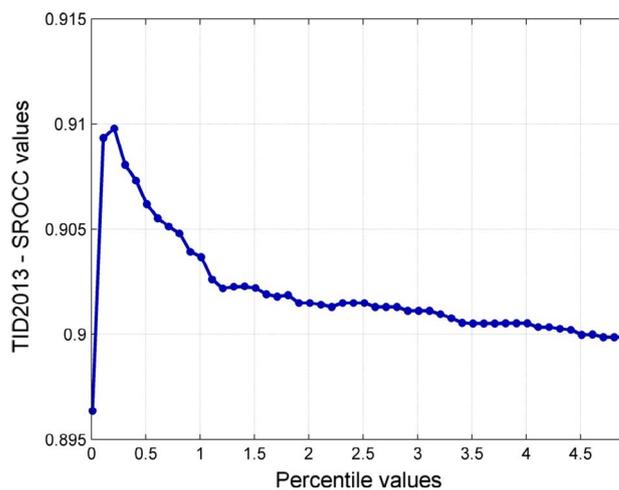[3] A MATLAB code computing the BISHARP objective scores is available for download from the Video Communication Laboratory website at http://www.vcl.fer.hr/quality/.
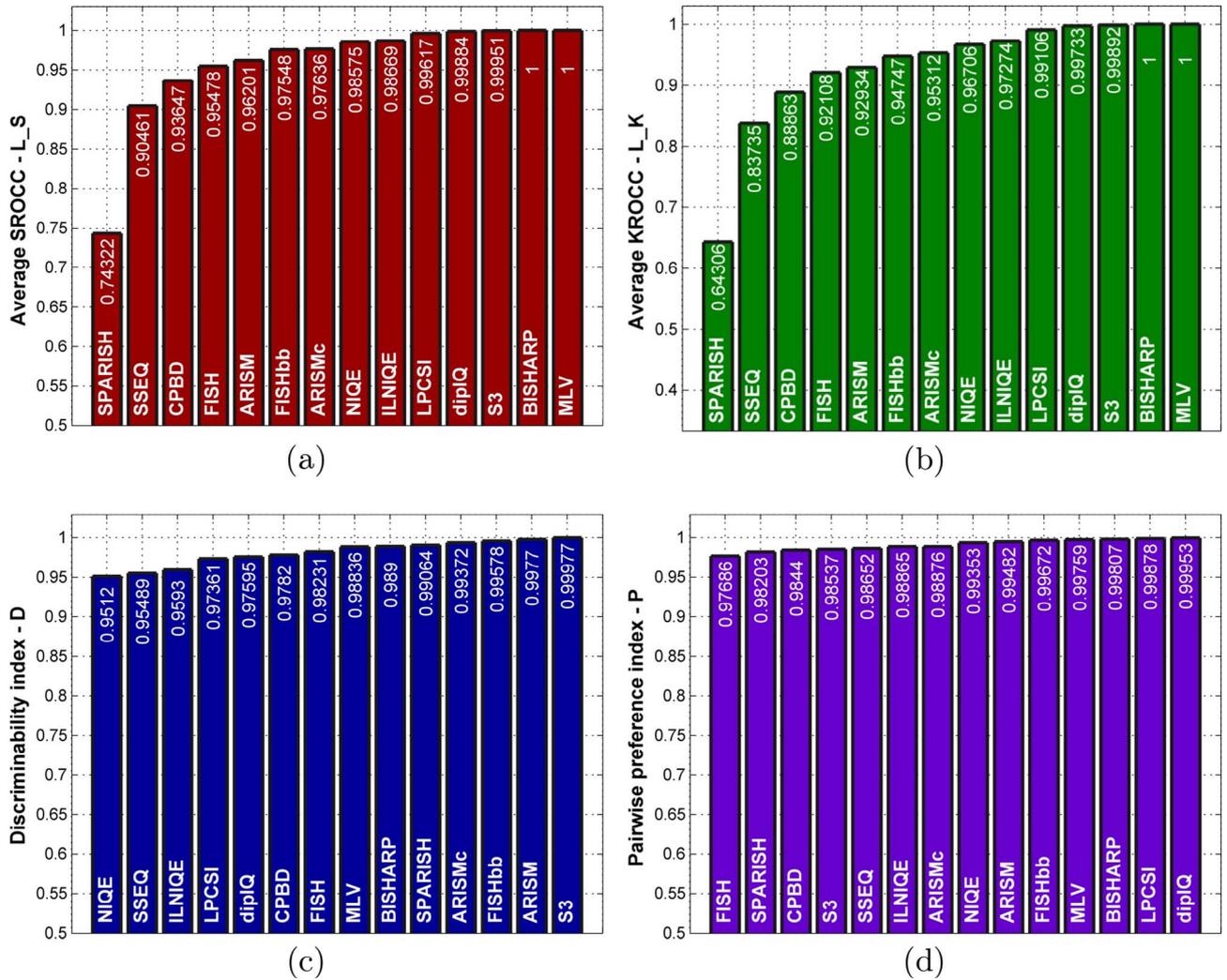
Fig. 10. The results of Listwise ranking consistency test based on average SROCC (a) and average KROCC (b), Discriminability test (c) and Pairwise preference test (d).

values obtained by BISHARP model fall in the mid results range. In case of IVC database, the MLV metric is competitive to BISHARP. Both methods reached very high correlation coefficients – PLCC/SROCC value pair for BISHARP is 0.9776/0.9797, while for MLV the value pair is 0.9777/0.9767. It should be noted that the IVC database – although small in size – represents a demanding database for existing image sharpness metrics. For example, ARISM and dipIQ methods were not able to keep the high performance level when tested on IVC. In addition, as a visual representation of performance evaluation the scatter plots of objective vs subjective scores after non-linear regression are provided in Fig. 6.

Moreover, to provide an overall performance comparison, we compute the weighted PLCC, SROCC and KROCC performance metrics for all databases. Hence, the overall correlation coefficients were computed as

$$CC_w = \frac{\sum_{i=1}^{N} w_i CC_i}{\sum_{i=1}^{N} w_i},$$ (21)

where CC_i represents the correlation score for the i-th database, N is the total number of tested databases and the weighting factors were determined by number of blurred images in each image database. Weighting factors, $w_i = \{174,138,20,150,100,125\}$, correspond to databases lined up in the following order: LIVE – 174, VCL@FER – 138, IVC – 20, CSIQ – 150, TID2008 – 100, TID2013 – 125. Besides weighted average we also compute the direct average for overall performance

comparison. The direct average is computed as the mean value of results obtained on all databases. The results of overall performance are shown in Table 2.

From Table 2 we can conclude that the proposed BISHARP method outperforms other state-of-the-art methods with the exception of weighted KROCC result obtained by SPARISH and weighted SROCC obtained by dipIQ. Hence, BISHARP is highly accurate, monotonic and consistent in sharpness estimation across different databases. In case of direct average results the BISHARP method turned out to be the best performing sharpness metric. In terms of overall performance comparison, the SPARISH, ARISM$_c$, ARISM and dipIQ methods highly correlate with human scores; however at the expense of increased computational complexity, as can be seen in Table 3.

3.4. Influence of parameters on performance metrics

The results of performance evaluation were affected by selected DWT level and sub-band, size of local pixel neighborhood, exponent α and percentile values. Hence, the model performance was observed at different DWT sub-bands while increasing the size of local window size, dynamic range and percentile values. As shown, the best alignment with the results of subjective measurement was obtained for BISHARP at 3-by-3 window size, 0.25th percentile and exponent $\alpha = 3.75$. This combination of local neighborhood size, pth percentile and dynamic range produced the best SROCC values and thus, was integrated within the proposed image sharpness assessment model. The SROCC value was
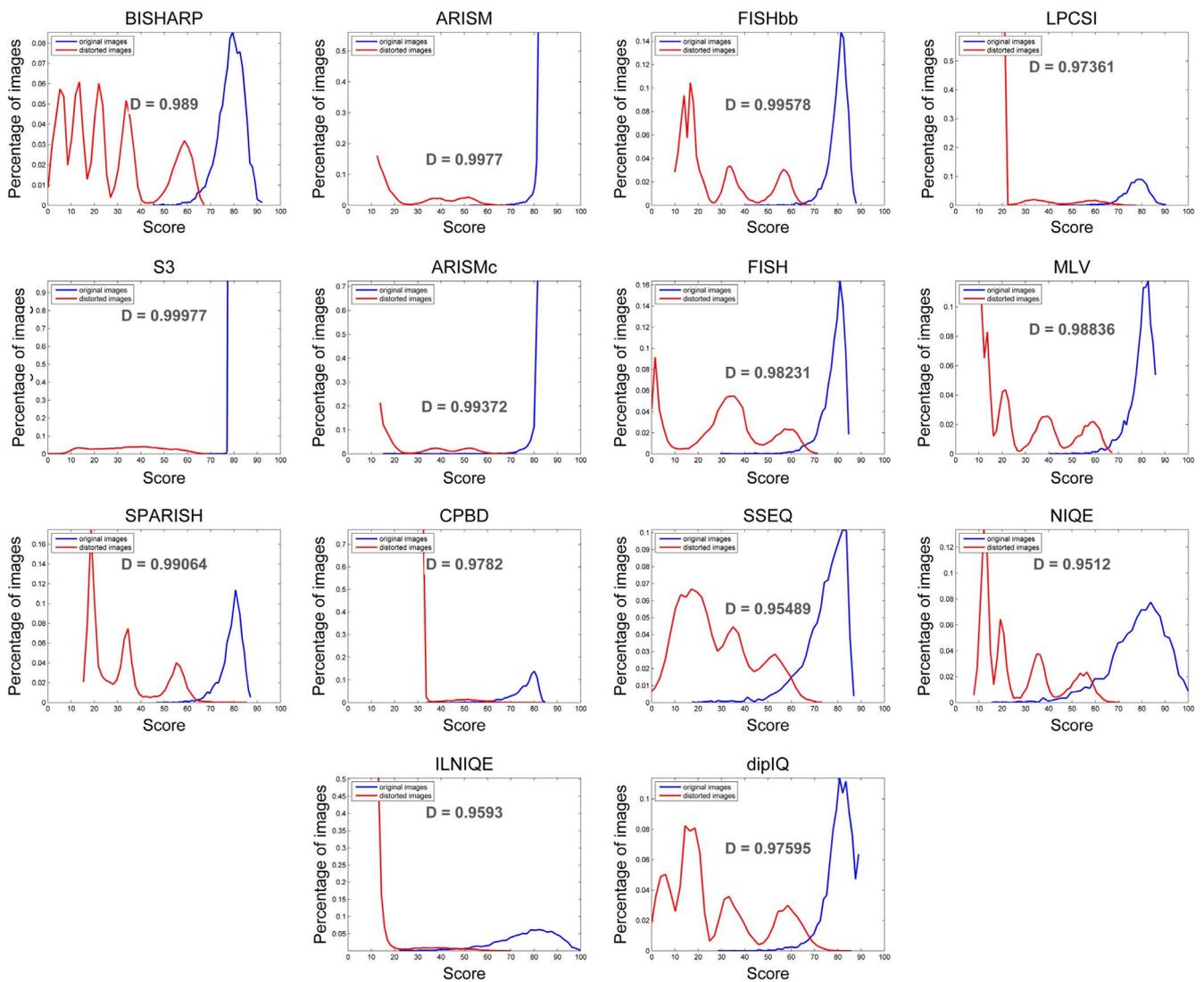
Fig. 11. Distributions of sharpness scores between original (in blue) and distorted images (in red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

used for comparison because it describes the relationship between the two data sets before non-linear regression. Thus, the influence of different data fitting functions was avoided. First, we observe the influence of DWT sub-bands on performance metrics.

3.4.1. Impact of DWT sub-band

To observe the impact of different wavelet sub-bands on performance metrics the three-scale DWT transform was performed to decompose the local contrast maps. As a result of decomposition we obtained nine high frequency sub-bands (horizontal – W_{H1}, W_{H2}, W_{H3}; vertical – W_{V1}, W_{V2}, W_{V3}; diagonal – W_{D1}, W_{D2}, W_{D3}) and a low frequency sub-band at third level (W_{A3}).

For each of the sub-bands including the low frequency sub-band the SROCC value was computed at 3-by-3 window size, $\alpha = 3.75$ and 0.25th percentile. The SROCC values were computed for all six conventional databases. The results of analysis for BISHARP are shown in Fig. 7. We can see that the performance decreases with higher decomposition levels; which is expected due to loss of high frequency components at higher decomposition levels.

3.4.2. Impact of window size

Influence of local window size on model's performance was tested using four different window sizes defined in 3-by-3, 5-by-5, 7-by-7 and 9-

by-9 pixel neighborhood. Fig. 8 depicts the influence of varying window size on performance metrics obtained on IVC database using BISHARP. As can be seen the local contrast maps generated using 3-by-3 window sizes produced the best correlation results. The results were obtained on the IVC database at 0.25th percentile and varying dynamic range. The performance falls significantly when using larger window sizes.

3.4.3. Impact of dynamic range

Increasing the dynamic range improves the performance metrics making the model's prediction capabilities more consistent and accurate. The influence of higher dynamic range on the model performance was tested by increasing the exponent α, in steps of $\alpha = \{1, 2, 3 \dots, n\}$. The results are shown in Fig. 8.

3.4.4. Impact of percentile values

In order to analyze how percentile values affect the model performance we computed the SROCC values for the complete percentile range in steps of 0.05 percentiles. Fig. 9 depicts the SROCC values as a function of percentile values. However, we show only a part of the percentile range – up to 5th percentile. The highest SROCC values were obtained in the lower percentile range. Higher percentile values reduced the model performance; however, not as much as above analyzed window sizes and dynamic range.
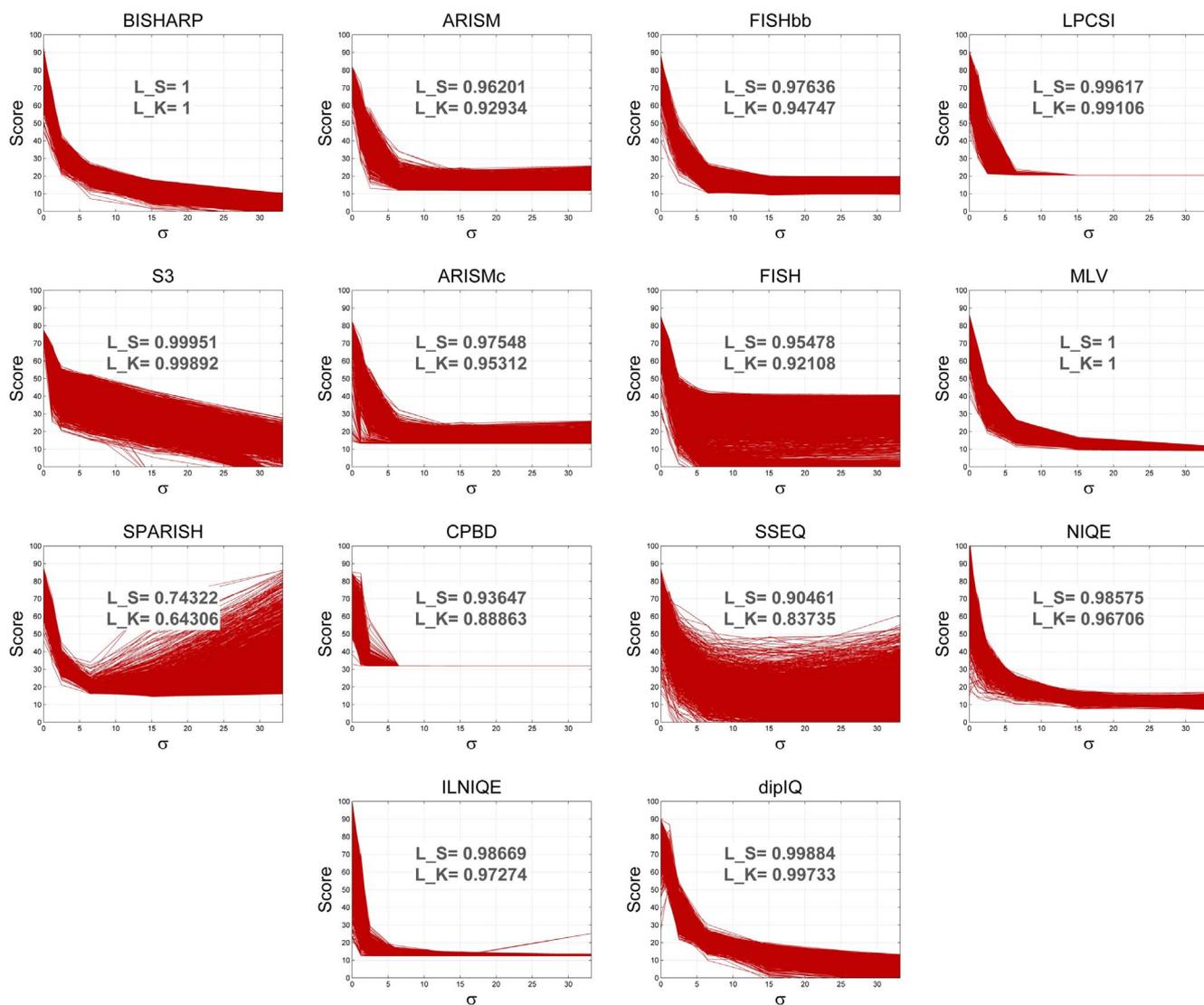
Fig. 12. Predicted scores as a function of increasing blur distortion expressed in σ values.

3.5. Results of performance evaluation on the Waterloo database

The results of tests performed on Waterloo database are shown in Fig. 10. First, the D-test was performed. The top performing methods in D-test are S3, both ARISM versions and block based FISH, as shown in Fig. 10(c). Our model follows these top performing methods together with SPARISH and MLV. The lowest ability to discriminate original from distorted images is demonstrated by general-purpose methods. This behavior is illustrated in Fig. 11 by plotting the distributions of sharpness scores for original and distorted images. The models with strong discriminability such as S3 or ARISM have small overlaps between original and distorted image distributions, while models with weak separability have visible overlaps between two distributions, as demonstrated by general-purpose methods. Moreover, the illustration shows that the top performing methods S3 and ARISM have lower discriminability at higher distortion levels. Good discriminability is manifested by BISHARP model where all five distortion levels can be clearly seen and differentiated. The similar behavior is demonstrated by MLV, SPARISH and FISH$_{bb}$ as well as by general-purpose, dipIQ and NIQE; however, with fewer number of visible peaks/distortion levels.

BISHARP is among the top performing models in the case of L-test, as shown in Fig. 10(a) and (b). Along with MLV, our method obtained the highest possible L_S and L_K values. The prediction monotonicity behavior is illustrated in Fig. 12 by plotting the sharpness scores as a function of increasing blur strengths expressed in terms of standard deviation $\sigma = \{1.2, 2.5, 6.5, 15.2, 33.2\}$. Hence, the sharpness scores for each image and its distorted versions in the blur subcategory of the Waterloo database were plotted against increasing sigma values. The plot for each method is made of 4744 curves/image sets. The similar approach has been taken by authors in [64,65,32,30] but with substantially fewer number of image sets. A good objective model should be able to produce sharpness scores that monotonically follow the increasing distortion levels. The smaller variation among the predicted scores at the same blur levels combined with good monotonic prediction ability reflects the better model. In that sense, we can see that the top performing methods, MLV and BISHARP show a small variation among curves. Methods such as CPBD, IL-NIQE and LPC-SI show inefficiency in discriminating severely blurred images, whereas ARISM and FISH – although highly aligned with perceptual scores – tend to produce ambiguous scores at higher blur distortion levels. The similar behavior is demonstrated by SSEQ and SPARISH where sharpness estimation ability at higher blur levels is significantly reduced and compromised.

The P-test was performed on more than 100 million DIPs whose sharpness was clearly discriminable. The total number of DIPs was extracted from the blur subcategory of the Waterloo database using quality scores obtained by MS-SSIM [36] and predicted MOS differences greater than threshold $T = 40$, as determined in [59]. The highest

number of correct preference predictions was obtained by dipIQ whereas LPC-SI and BISAHRP methods showed to be very competitive by achieving slightly inferior results, as shown in Fig. 10(d). The results of performed tests confirmed those obtained on conventional databases. Hence, the BISHARP model based on local contrast maps proved its prediction monotonicity and prediction consistency.

3.6. Computational complexity

A new blurred image database was created out of images found in six available databases. The new database, used only for computational complexity evaluation was made out of 39 images from LIVE database, 24 from CISQ, 20 from VCL@FER and 17 from TID2013 database. The image sizes range from 512×512, 480×720, 634×505, 683×512 to 768×512. The experiments were performed on a notebook with Intel Core i5 CPU at 1.70 GHz, 4 GB RAM, 64-bit operating system and Matlab. No additional efforts or optimization techniques were used to obtain the results in Table 3. The best run time was obtained by FISH method (0.11 s per image), whereas the second best result, with somewhat slower run time per image, was obtained by our BISHARP model (0.13 s per image). The third best result is obtained by spatial domain based MLV metric (0.17 s per image). The ARISM, $ARISM_c$ and SPARISH objective models characterized with very high correlation results have shown to be very time consuming – run time range is from 15.32 s/image for learning based SPARISH to 66.76 s/image for $ARISM_c$. The hybrid spatio-frequency based S3 metric has shown the worst result in this run time test.

3.7. Cross validation based on statistical Significance Analysis

The performance results obtained by model being evaluated are considered statistically significant if it is unlikely that the results were obtained by chance. By employing the statistical significance testing we are able to verify the objective model and relate it to other models found in the image sharpness/blurriness ecosystem. Hence, hypothesis test is conducted to determine whether the performance of the proposed BISHARP is statistically better, indistinguishable or worse in comparison to other considered methods. Therefore, to get a better picture of the current image sharpness/blurriness assessment research ecosystem a state-of-the-art methods were observed and compared.

Statistical significance testing was performed using the F-test where the variance ratio of two data sets was observed. In our case data sets are represented by absolute values remained after subtraction between subjective scores (y) and predicted scores (y_p) obtained after nonlinear regression. Residuals are computed according to the following expression

$$e_{metric} = |y - y_p|, \tag{22}$$

where e_{metric} represents residuals of method being under observation. With F-test the variance ratio or F-statistic is computed and observed within the 95% confidence interval or at 5% significance level. We hypothesize that residuals come from normal distributions with the same variance. If that is the case the null hypothesis is accepted meaning that the considered metrics are statistically indistinguishable. The alternative hypothesis is that the variance of first metric is lower or higher than the variance of the second metric. Hence, the left tailed F-test is employed to test if the residual variance of first metric is lower than the variance of the second metric whereas the right tailed F-test is employed to test the opposite behavior. Results of statistical significance testing are presented in Table 4. A symbol 0 indicates that the row method is statistically indistinguishable to the column method. A symbol 1 denotes that the row method is statistically better than the column method, whereas a symbol -1 denotes that the row method is statistically worse than the column method. In comparison to eight tested methods, namely SSEQ, NIQE, IL-NIQE, LPC-SI, FISH, $FISH_{bb}$, S3 and MLV the F-test showed that BISHARP is statistically better or

indistinguishable on all six databases. The similar behavior was demonstrated in comparison to remaining methods except in the case of VCL@FER database where BISHARP showed statistical inferiority to SPARISH, dipIQ, CPBD and both ARISM versions.

4. Concluding remarks and future work

In this paper we presented the importance of local contrast measurement in estimation of image sharpness/blurriness. We observed how local contrast map elements behave in the presence of increasing blur distortion both, in the spatial and frequency domain. Along with the performed analysis we presented the framework utilized to identify and extract the image sharpness feature. Hence, the percentile value of ranked, level-shifted high-frequency wavelet coefficients was computed, extracted and defined as a perceptually significant image sharpness feature capable to accurately estimate image sharpness. A fast, simple and computationally efficient algorithm was developed and validated on seven publicly available databases containing blurred images. A thorough performance evaluation demonstrated very high correlation with the perceptual scores. Moreover, the proposed BISHARP model was proven to be very consistent in sharpness estimation across all tested databases. These findings combined with low computational complexity make the BISHARP model highly efficient and effective in estimating the image sharpness.

In the scope of future work an analysis of local contrast statistics in a multi-scale and multi-orientation transform space can be performed. Applying BISHARP within the image segmentation research area is worth exploring since the sharpness maps generated at lower dynamic range represent a good foundation to differentiate the foreground/sharp and background/blurred image areas. Furthermore, it would be interesting to see how the extracted sharpness feature can be used in a holistic IQA framework based on machine learning.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jvcir.2017.11.017.

References

[1] M.A. Webster, M.A. Georgeson, S.M. Webster, Neural adjustments to image blur, Nat. Neurosci. 5 (9) (2002) 839–840.

[2] A.K. Jain, Fundamentals of Digital Image Processing, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1989.

[3] S. Hillaire, A. Lécuyer, R. Cozot, G. Casiez, Using an eye-tracking system to improve camera motions and depth-of-field blur effects in virtual environments, in: IEEE Virtual Reality Conference 2008 (VR 2008), 8–12 March 2008, Reno, Nevada, USA, Proceedings, 2008, pp. 47–50, http://dx.doi.org/10.1109/VR.2008.4480749.

[4] J. Shi, L. Xu, J. Jia, Discriminative blur detection features, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 2965–2972.

[5] K. Ma, H. Fu, T. Liu, Z. Wang, D. Tao, Local Blur Mapping: Exploiting High-level Semantics by Deep Neural Networks. Available from: < arXiv:1612.01227 > .

[6] J. Shi, L. Xu, J. Jia, Just noticeable defocus blur detection and estimation, in: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7–12, 2015, pp. 657–665, http://dx.doi.org/10.1109/CVPR.2015.7298665.

[7] J. Guan, W. Zhang, J.J. Gu, H. Ren, No-reference blur assessment based on edge modeling, J. Visual Commun. Image Rep. 29 (2015) 1–7, http://dx.doi.org/10.1016/j.jvcir.2015.01.007.

[8] X. Zhang, R. Wang, X. Jiang, W. Wang, W. Gao, Spatially variant defocus blur map estimation and deblurring from a single image, J. Visual Commun. Image Rep. 35 (2016) 257–264, http://dx.doi.org/10.1016/j.jvcir.2016.01.002.

[9] ITU-T P.910, Recommendation, Subjective Video Quality Assessment Methods for Multimedia Applications, 2006.

[10] S. Xu, S. Jiang, W. Min, No-reference/blind image quality assessment: a survey, IETE Tech. Rev. 34 (3) (2017) 223–245.

[11] Z. Wang, A.C. Bovik, H.R. Sheikh, Structural similarity based image quality assessment, Digital Video Image Qual. Percept. Coding (2005) 225–241.

[12] M. Shahid, A. Rossholm, B. Lövström, H.-J. Zepernick, No-reference image and video quality assessment: a classification and review of recent approaches, EURASIP J. Image Video Process. 2014 (1) (2014) 40.

[13] P. Marziliano, F. Dufaux, S. Winkler, T. Ebrahimi, A no-reference perceptual blur

metric, in: Proceedings of the 2002 International Conference on Image Processing, ICIP 2002, Rochester, New York, USA, September 22–25, 2002, pp. 57–60, http://dx.doi.org/10.1109/ICIP.2002.1038902.

[14] R. Ferzli, L.J. Karam, A no-reference objective image sharpness metric based on just-noticeable blur and probability summation, in: Proceedings of the International Conference on Image Processing, ICIP 2007, September 16–19, 2007, San Antonio, Texas, USA, 2007, pp. 445–448, http://dx.doi.org/10.1109/ICIP.2007.4379342.

[15] N.D. Narvekar, L.J. Karam, A no-reference image blur metric based on the cumulative probability of blur detection (CPBD), IEEE Trans. Image Process. 20 (9) (2011) 2678–2683.

[16] K. Bahrami, A.C. Kot, A fast approach for no-reference image sharpness assessment based on maximum local variation, IEEE Signal Process. Lett. 21 (6) (2014) 751–755, http://dx.doi.org/10.1109/LSP.2014.2314487.

[17] K. Gu, G. Zhai, W. Lin, X. Yang, W. Zhang, No-reference image sharpness assessment in autoregressive parameter space, IEEE Trans. Image Process. 24 (10) (2015) 3218–3231, http://dx.doi.org/10.1109/TIP.2015.2439035.

[18] L. Kang, P. Ye, L. Yi, D. Doermann, Convolutional neural networks for no-reference image quality assessment, Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR '14, IEEE Computer Society, Washington, DC, USA, 2014, pp. 1733–1740, , http://dx.doi.org/10.1109/CVPR.2014.224.

[19] K. Ma, W. Liu, T. Liu, Z. Wang, D. Tao, dipIQ: Blind image quality assessment by learning-to-rank discriminable image pairs, IEEE Trans. Image Process. 26 (8) (2017) 3951–3964, http://dx.doi.org/10.1109/TIP.2017.2708503.

[20] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, IEEE Trans. Image Process. 21 (12) (2012) 4695–4708, http://dx.doi.org/10.1109/TIP.2012.2214050.

[21] L. Li, D. Wu, J. Wu, H. Li, W. Lin, A.C. Kot, Image sharpness assessment by sparse representation, IEEE Trans. Multimedia 18 (6) (2016) 1085–1097, http://dx.doi.org/10.1109/TMM.2016.2545398.

[22] L. Li, W. Lin, X. Wang, G. Yang, K. Bahrami, A.C. Kot, No-reference image blur assessment based on discrete orthogonal moments, IEEE Trans. Cybernet. 46 (1) (2016) 39–50, http://dx.doi.org/10.1109/TCYB.2015.2392129.

[23] Q.-B. Sang, H.-X. Qi, X.-J. Wu, C.-F. Li, A.C. Bovik, No-reference image blur index based on singular value curve, J. Visual Commun. Image Rep. 25 (7) (2014) 1625–1630, http://dx.doi.org/10.1016/j.jvcir.2014.08.002.

[24] Q.-B. Sang, X.-J. Wu, C.-F. Li, Y. Lu, Blind image blur assessment using singular value similarity and blur comparisons, PLOS ONE 9 (9) (2014) 1–6, http://dx.doi.org/10.1371/journal.pone.0108073.

[25] J. Caviedes, F. Oberti, A new sharpness metric based on local kurtosis, edge and energy information, Signal Process.: Image Commun. 19 (2) (2004) 147–161.

[26] M.A. Saad, A.C. Bovik, C. Charrier, DCT statistics model-based blind image quality assessment, in: 18th IEEE International Conference on Image Processing, ICIP 2011, Brussels, Belgium, September 11–14, 2011, pp. 3093–3096, http://dx.doi.org/10.1109/ICIP.2011.6116319.

[27] A.K. Moorthy, A.C. Bovik, Blind image quality assessment: from natural scene statistics to perceptual quality, IEEE Trans. Image Process. 20 (12) (2011) 3350–3364, http://dx.doi.org/10.1109/TIP.2011.2147325.

[28] Z. Wang, E.P. Simoncelli, Local phase coherence and the perception of blur, in: Advances in Neural Information Processing Systems 16, NIPS 2003, December 8–13, 2003, Vancouver and Whistler, British Columbia, Canada, 2003, pp. 1435–1442.

[29] R. Hassen, Z. Wang, M. Salama, No-reference image sharpness assessment based on local phase coherence measurement, in: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2010, 14–19 March 2010, Dallas, Texas, USA, 2010, pp. 2434–2437, http://dx.doi.org/10.1109/ICASSP.2010.5496297.

[30] R. Hassen, Z. Wang, M. Salama, Image sharpness assessment based on local phase coherence, IEEE Trans. Image Process. 22 (7) (2013) 2798–2810, http://dx.doi.org/10.1109/TIP.2013.2251643.

[31] P.V. Vu, D.M. Chandler, A fast wavelet-based algorithm for global and local image sharpness estimation, IEEE Signal Process. Lett. 19 (7) (2012) 423–426, http://dx.doi.org/10.1109/LSP.2012.2199980.

[32] C.T. Vu, T.D. Phan, D.M. Chandler, S3: a spectral and spatial measure of local perceived sharpness in natural images, IEEE Trans. Image Process. 21 (3) (2012) 934–945, http://dx.doi.org/10.1109/TIP.2011.2169974.

[33] L. Li, W. Xia, W. Lin, Y. Fang, S. Wang, No-reference and robust image sharpness evaluation based on multiscale spatial and spectral features, IEEE Trans. Multimedia. 19 (5) (2017) 1030–1040, http://dx.doi.org/10.1109/TMM.2016.2640762.

[34] L. Zhang, L. Zhang, A.C. Bovik, A feature-enriched completely blind image quality evaluator, IEEE Trans. Image Process. 24 (8) (2015) 2579–2591.

[35] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612, http://dx.doi.org/10.1109/TIP.2003.819861.

[36] Z. Wang, E.P. Simoncelli, A.C. Bovik, Multi-scale structural similarity for image quality assessment, in: Proceedings of the IEEE Conference on Signals, Systems, and Computers, Asilomar, 2003, pp. 1398–1402.

[37] W. Xue, X. Mou, L. Zhang, A.C. Bovik, X. Feng, Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features, IEEE Trans. Image Process. 23 (11) (2014) 4850–4862, http://dx.doi.org/10.1109/TIP.2014.2355716.

[38] M.A. Saad, A.C. Bovik, C. Charrier, A DCT statistics-based blind image quality index, IEEE Signal Process. Lett. 17 (6) (2010) 583–586, http://dx.doi.org/10.1109/LSP.2010.2045550.

[39] S. Winkler, Visual fidelity and perceived quality: toward comprehensive metrics, in: Human Vision and Electronic Imaging VI, San Jose, CA, USA, January 20, 2001, pp. 114–125, http://dx.doi.org/10.1117/12.429540.

[40] E. Peli, Contrast in complex images, J. Opt. Soc. Am. 7 (1990) 2032–2040.

[41] R.A. Frazor, W.S. Geisler, Local luminance and contrast in natural images, Vision. Res. 46 (10) (2006) 1585–1598, http://dx.doi.org/10.1016/j.visres.2005.06.038.

[42] D.C. Knill, D. Field, D. Kerstent, Human discrimination of fractal images, J. Opt. Soc. Am. A 7 (6) (1990) 1113–1123, http://dx.doi.org/10.1364/JOSAA.7.001113.

[43] R. Mantiuk, High dynamic range imaging: towards the limits of the human visual perception, Forsch. Wiss. Rechnen 72 (2007) 11–27.

[44] C. Poynton, Digital Video and HDTV Algorithms and Interfaces, first ed., Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2003.

[45] Y. Gao, A. Rehman, Z. Wang, CW-SSIM based image classification, in: 18th IEEE International Conference on Image Processing, ICIP 2011, Brussels, Belgium, September 11–14, 2011, pp. 1249–1252, http://dx.doi.org/10.1109/ICIP.2011.6115659.

[46] ITU-R BT.601-7, Recommendation, Studio Encoding Parameters of Digital Television for Standard 4:3 and Wide-screen 16:9 Aspect Ratios, March 2011.

[47] The MathWorks Inc., Matlab R2013a, February 2013.

[48] Z. Wang, X. Shang, Spatial pooling strategies for perceptual image quality assessment, in: Proceedings of the International Conference on Image Processing, ICIP 2006, October 8–11, Atlanta, Georgia, USA, 2006, pp. 2945–2948, http://dx.doi.org/10.1109/ICIP.2006.313136.

[49] P.J. Bex, W. Makous, Spatial frequency, phase, and the contrast of natural images, J. Opt. Soc. Am. A 19 (6) (2002) 1096–1106, http://dx.doi.org/10.1364/JOSAA.19.001096.

[50] D.G. Pelli, P. Bex, Measuring contrast sensitivity, Vision. Res. 90 (2013) 10–14, http://dx.doi.org/10.1016/j.visres.2013.04.015 (testing Vision: From Laboratory Psychophysical Tests to Clinical Evaluation).

[51] E.P. Simoncelli, E.H. Adelson, Subband transforms, in: J.W. Woods (Ed.), Subband Image Coding, Kluwer Academic Publishers, Norwell, MA, 1990, pp. 143–192 (Chapter 4).

[52] H. Sheikh, Z. Wang, L. Cormack, A. Bovik, Live Image Quality Assessment Database Release, vol. 2, 2005.

[53] M.S.H.R. Sheikh, A. Bovik, A statistical evaluation of recent full reference image quality assessment algorithms, IEEE Trans. Image Process. 15 (11) (2006) 3440–3451, http://dx.doi.org/10.1109/TIP.2006.881959.

[54] E.C. Larson, D.M. Chandler, Most apparent distortion: full-reference image quality assessment and the role of strategy, J. Electron. Imaging 19 (1) (2010) 011006, http://dx.doi.org/10.1117/1.3267105.

[55] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, F. Battisti, TID2008—a database for evaluation of full-reference visual quality assessment metrics, Adv. Modern Radioelectron. 10 (4) (2009) 30–45.

[56] N.N. Ponomarenko, O. Ieremeiev, V.V. Lukin, K.O. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, C.J. Kuo, Color image database TID2013: peculiarities and preliminary results, in: European Workshop on Visual Information Processing, EUVIP 2013, Paris, France, June 10–12, 2013, pp. 106–111.

[57] A. Zaric, N. Tatalovic, N. Brajkovic, H. Hlevnjak, M. Loncaric, E. Dumic, S. Grgic, VCL@FER image quality assessment database, Automatika 53 (4) (2012) 344–354.

[58] P.L. Callet, F. Autrusseau, Subjective Quality Assessment IRCCyN/IVC Database, 2005.

[59] K. Ma, Z. Duanmu, Q. Wu, Z. Wang, H. Yong, H. Li, L. Zhang, Waterloo exploration database: new challenges for image quality assessment models, IEEE Trans. Image Process. 26 (2) (2017) 1004–1016, http://dx.doi.org/10.1109/TIP.2016.2631888.

[60] ITU-T Tutorial, Objective Perceptual Assessment of Video Quality: Full Reference Television, ITU-T Telecommunication Standardization Bureau, 2004.

[61] N.D. Narvekar, L.J. Karam, A no-reference image blur metric based on the cumulative probability of blur detection, IEEE Trans. Image Process. 20 (9) (2011) 2678–2683, http://dx.doi.org/10.1109/TIP.2011.2131660.

[62] L. Liu, B. Liu, H. Huang, A.C. Bovik, No-reference image quality assessment based on spatial and spectral entropies, Sig. Proc.: Image Commun. 29 (8) (2014) 856–863, http://dx.doi.org/10.1016/j.image.2014.06.006.

[63] A. Mittal, R. Soundararajan, A.C. Bovik, Making a "completely blind" image quality analyzer, IEEE Signal Process. Lett. 20 (3) (2013) 209–212.

[64] R. Ferzli, L.J. Karam, Human visual system based no-reference objective image sharpness metric, in: Proceedings of the International Conference on Image Processing, ICIP 2006, October 8–11, Atlanta, Georgia, USA, 2006, pp. 2949–2952, http://dx.doi.org/10.1109/ICIP.2006.312925.

[65] P. Marziliano, F. Dufaux, S. Winkler, T. Ebrahimi, A no-reference perceptual blur metric, in: Proceedings of the 2002 International Conference on Image Processing, ICIP 2002, Rochester, New York, USA, September 22–25, 2002, pp. 57–60, http://dx.doi.org/10.1109/ICIP.2002.1038902.