J. Vis. Commun. Image R. 60 (2019) 140–148

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Reduced-reference quality assessment of image super-resolution by energy change and texture variation ☆

Yuming Fang [a], Jiaying Liu [b,*], Yabin Zhang [c], Weisi Lin [c], Zongming Guo [b]

[a] School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, Jiangxi 330032, China
[b] Institute of Computer Science and Technology, Peking University, Beijing 100080, China
[c] School of Computer Science and Engineering, Nanyang Technological University, 639798, Singapore

ARTICLE INFO

Article history:
Received 28 November 2017
Revised 19 August 2018
Accepted 18 December 2018
Available online 2 February 2019

Keywords:
Image quality assessment (IQA)
Image super-resolution
Reduced-reference (RR) quality assessment
Energy change
Texture variation

ABSTRACT

In this paper, we propose a novel reduced-reference quality assessment metric for image super-resolution (RRIQA-SR) based on the low-resolution (LR) image information. With the pixel correspondence, we predict the perceptual similarity between image patches of LR and SR images by two components: the energy change in low-frequency regions, which can be used to capture the global distortion in SR images, and texture variation in high-frequency regions, which can be used to capture the local distortion in SR images. The overall quality of SR images is estimated by perceptual similarity calculated by energy change and texture variation between local image patches of LR and HR images. Experimental results demonstrate that the proposed method can obtain better performance of quality prediction for SR images than other existing ones, even including some full-reference (FR) metrics.

1. Introduction

Image super-resolution (SR) technique provides an effective solution for the problem of image resolution limitation from some specific imaging sensors such as surveillance cameras, mobile devices, etc. With image super-resolution, the low-resolution (LR) images from these devices can be better displayed and utilized in the common high-resolution (HR) displays and thus provide good visual experiences for users. In many multimedia processing systems, image super-resolution is highly desired to obtain SR images from LR images for some specific tasks such as detection, recognition, etc. During the past decades, there have been numerous image super-resolution algorithms proposed for various multimedia processing applications [1–15], and there have been various applications for image super-resolution including medical image processing, infrared imaging, face/iris recognition, image editing, virtual reality (VR), *etc.*

Early methods use shift and aliasing properties of Fourier transform for image super-resolution. These methods are efficient, but they cannot model the complicated image degradation and image priors. To overcome the drawbacks of the methods in frequency domain, various spatial image super-resolution methods have been designed. One simple spatial image super-resolution method is image interpolation, which tries to obtain HR images from LR images by pixel interpolation [5,36]. The problem with image interpolation is that there are serious aliasing artifacts and blurring distortions along edges and high-frequency regions due to the pixel interpolation operation.

To overcome these problems, many advanced image super-resolution algorithms have been proposed, including reconstruction-based methods, learning-based methods, *etc.* [1]. Reconstruction-based image super-resolution methods generate HR images by a regularized cost function with certain prior knowledge [6,37,11], while the example learning-based method reconstructs HR images by learning the mapping function between image patches from LR images to HR images. The exemplar image patches can be extracted from the input image, the external databases, or combined sources [8].

For these existing image super-resolution studies introduced above, the performance of image super-resolution algorithms is mainly validated by small-scale subjective tests. The problem with subjective tests is that they are time-consuming and require subjects involved in the experiments, and thus they cannot be used in practical systems. Currently, much less has been done to assess the visual quality of HR images objectively. Existing visual quality assessment metrics such as peak signal-to-noise-ratio (PSNR), structural similarity (SSIM) [18] and others cannot be used in

☆ This article is part of the Special Issue on Visual Info. Proc. For VR.

* Corresponding author.
E-mail addresses: fa0001ng@e.ntu.edu.sg (Y. Fang), liujiaying@pku.edu.cn (J. Liu), zhan0398@e.ntu.edu.sg (Y. Zhang), wslin@ntu.edu.sg (W. Lin), guozongming@pku.edu.cn (Z. Guo).

https://doi.org/10.1016/j.jvcir.2018.12.035
1047-3203/

super-resolution applications, since they can be used only in the cases where the sizes of the reference and distorted images are the same. Recently, there are only a few objective image quality assessment (IQA) studies investigating the visual quality assessment of HR images [20]. However, the study mainly focuses on quality assessment for interpolated natural images [20]. The performance of existing IQA metrics is low in visual quality assessment of image super-resolution, as demonstrated by the experimental results in Section 4. Thus, it is highly desired to design the effective objective quality assessment metric for image super-resolution for various practical systems of image super-resolution.

In this study, we propose a novel reduced-reference (RR) quality metric for image super-resolution (RRIQA-SR). For image super-resolution, the global structural information in the generated HR image should be generated based upon that in the LR image. For the generated structural information in the HR image, it can be divided into two parts: low-frequency and high-frequency regions. The visual quality of the HR image can be estimated by visual distortion in these two parts. To measure the visual quality in HR images, the energy and texture features are extracted for low-frequency and high-frequency regions of image patches in LR and HR images, respectively. The feature difference from energy and texture are used to estimate the perceptual similarity between LR and HR images, which is further adopted to predict the overall quality of HR images. Here, the used energy change and texture variation can capture the global and local distortions in HR images, respectively. Experimental results show that the proposed RRIQA-SR can obtain better performance in quality prediction of HR images than other existing ones. Please note that the initial work was published in the study [59].

The remaining of this paper is organized as follows. Section 2 introduces the related work of image super-resolution and visual quality assessment in the literature. Section 3 describes the proposed method in detail. In Section 4, we provide the experimental results from different quality metrics to demonstrate the performance of the proposed method. The final section concludes the paper.

2. Related work

2.1. Image super-resolution

Considering the number of available LR images, image super-resolution algorithms can be classified into two categories: multi-frame super-resolution and single-frame super-resolution approaches [9]. For the multi-frame super-resolution methods, it can be further divided into two groups: the static super-resolution approach which only uses the corresponding LR frames to generate the current HR frame [33,34], while the dynamic super-resolution approach which use previous reconstructed HR frames to obtain the current HR frame [35].

For single-image super-resolution methods, there have been many different approaches proposed previously. Traditional interpolation based image super-resolution methods try to reconstruct the HR image by a base function, including bilinear, bicubic and nearest neighbor algorithms [5,36,39]. In the study [39], Li et al. proposed an edge-directed interpolation algorithm for natural images based on bilinear interpolation and covariance-based adaptive interpolation. An edge-guided image interpolation method was designed by Zhang et al. based on directional filtering and data fusion [40]. Recently, Wei et al. adopted contrast information to design image interpolation algorithm [36]. As introduced previously, the image interpolation is generally simple and efficient, but they suffer from artifacts/distortions easily in high-frequency regions.

In the past years, there were many reconstruction-based image super-resolution methods proposed by prior information to try to obtain better performance than traditional image interpolation algorithms. In the study [6], the authors used the statistical edge features to reconstruct HR images by resolving a constrained optimization problem. Sun et al. adopted gradient profile prior describing the sharpness of the image gradient to reconstruct the HR images [7]. In the study [37], the self-similarities of natural images were used as the prior for image super-resolution. Recently, sparsity prior has been widely used as prior information in image super-resolution [10,11,42], *etc.*. In these reconstruction-based image super-resolution methods, the used prior information is extracted based on some properties of natural images, and thus some blurring distortion or aliasing artifacts can be suppressed in HR images. However, the prior information used in the reconstruction-based methods would lead to distorted fine image structures if the up-scaling factor is large.

Recently, a popular type of image super-resolution methods is the example learning-based method, which reconstructs HR images by learning the mapping function between image patches from LR images to HR images [14,38,41,15]. In the study [12], support vector regression (SVR) was used to learn the mapping function in DCT domain between LR and HR images. Yang et al. used sparse dictionary representation to learn the mapping function between LR and HR images [14]. Following this work, there have been various sparse dictionary representation methods proposed for image super-resolution [9,13]. Zhang et al. used clustering and collaborative representation to propose a image super-resolution algorithm by learning the statistical priors [54]. The deep learning technique was also used in a recent study [15], *etc.* to learn the mapping function between LR and HR images for image super-resolution.

2.2. Image quality assessment

There are two types of visual quality assessment methods: subjective quality assessment and objective quality assessment. Since the human visual system (HVS) is the final receiver for visual content, subjective quality assessment is an accurate and reliable method for image quality assessment. During subjective experiments, a number of observers have to be invited to participate in the test to provide a rating score for each image. The average score overall all subjects, considered as mean opinion score (MOS), is used to represent the subjective score for each image. Although subjective quality assessment can obtain accurate and robust quality prediction for visual content, it is expensive, time-consuming, and cannot be embedded into super-resolution algorithms for optimization purpose [17,43].

To perform visual quality assessment in practical applications, there have been various objective quality metrics proposed to estimate visual quality of visual content consistent with human perception. According to the availability of the reference image, there are three types of image quality assessment (IQA) metrics [17,43]: full-reference (FR) metrics [18,29,19,44,55,57], RR metrics [47–49,56], and no-reference (NR) metrics [45,51,58]. The difference between these three types of IQA metrics is as follows: the FR metric requires the complete original image for visual quality prediction of the distorted image; the RR metric requires part information of the original image for visual quality estimation of the distorted image; the NR metric does not require any information of the original image for visual quality prediction of the distorted image. Generally, FR metrics can predict more accurate visual quality of distorted images compared with RR and NR metrics, since there is more available reference information.

During the past decades, there has been much progress in the area of objective IQA [16,17,43]. Traditional signal fidelity metrics such as PSNR, mean absolute error (MAE), mean square error (MSE), *etc.* predict visual quality of images by simply comparing the reference and distorted images without taking the visual content into account. These methods are simple and efficient, but they cannot estimate the visual quality of images accurately due to the lack of visual perception factors [16,17]. To better predict the visual quality of visual content, there have been various perceptual IQA metrics proposed during the past decade [17,43,47].

In the study [18], Wang et al. proposed the structural similarity metric (SSIM) by considering the perceptual characteristics of visual structure in images. SSIM has received much attention and been widely used in various multimedia systems in the past years. Later, Sheikh et al. designed a IQA metric called visual information fidelity (VIF) based on random field from the subband [29]. Larson et al. used visual masking and local statistics of spatial frequency components to devise a perceptual IQA metric of most apparent distortion [30]. In the study [19], the authors adopted the concept of internal generative mechanism (IGM) for IQA. There are many other IQA metrics proposed based on gradient similarity [44], contrast features [46], free energy [50], and so on.

As introduced previously, most existing perceptual IQA metrics cannot be used for image super-resolution, since they need the sizes of the reference and distorted images to be the same. For image super-resolution, the sizes of original LR and generated HR images are different. In the past, there were several studies investigating the visual quality assessment of image super-resolution subjectively and objectively [21,20,8]. Reibman et al. conducted subjective tests to evaluate the visual quality of super-resolution enhanced images [21]. That study also demonstrates that even FR metrics such as SSIM cannot always capture visual quality of HR images [21]. The authors in [20] proposed an objective IQA metric based on natural scene statistics (NSS). However, that NSS based method is mainly designed for interpolated natural images [20].

Recently, Yang et al. conducted a subjective study for quality evaluation of single-frame super-resolution by using some state-of-the-art single-frame super-resolution methods [8]. The full-reference IQA metrics such as PSNR, SSIM, *etc.* are used to evaluate the visual quality of HR images. However, in most practical applications, the only available information is the LR image and there is no ground truth HR image. Thus, it is highly desirable to design IQA metrics for HR images with only available LR images or without any reference information.

2.3. Contributions of our work

In this study, we investigate the objective visual quality assessment for single-frame super-resolution and propose a RRIQA-SR metric; due to the use of only LR image information, it is an RR type of IQA because the LR image to start with can be regarded as partial reference to the generated SR image. In fact, RR IQA is the most meaningful and practical IQA for super-resolution construction. The proposed RRIQA-SR is designed based on the perceptual similarity between LR and HR images. The MRF is first used to model pixel correspondence between LR and HR images. Then the energy and texture features are extracted in the low-frequency and high-frequency regions in image patches of LR and HR images by DCT coefficients. The perceptual similarity between LR and HR images is calculated by the feature difference between image patches of LR and HR images. The main contributions of this study include the following aspects.

- To measure the overall visual degradation in HR images, we calculate the energy change from energy differences between image patches of LR and HR. The DC coefficients are used to extract the energy features of low-frequency regions in image patches. The designed energy change can capture the global distortion in HR images.
- To measure the detailed visual distortion in high-frequency regions of HR images, we compute the texture variation from the texture feature differences between image patches of LR and HR images. The AC coefficients are used to obtain the texture features of high-frequency regions in image patches. The designed texture variation can capture the local distortion in HR images.

3. Proposed method

In this section, we introduce the proposed method in detail. We analyze the procedure of image super-resolution and provide the framework of the proposed method in the first subsection. The process of pixel correspondence is then described. Following that, we give the computation of energy change and texture variation for visual quality prediction of image super-resolution. The visual quality prediction model is provided in the final subsection.

3.1. Overview

During image super-resolution construction, the overall visual information of the generated SR image should be highly similar with that of the original LR image. For the reconstructed SR images, the visual distortion brought into during image super-resolution operation is mainly caused from the following two aspects: one is the overall energy change of low-frequency regions from a LR image to its generated HR image, while the other is the visual artifacts in high-frequency regions such as edges, corners, *etc.* These two aspects can be also considered as global and local distortion in SR images.

In Fig. 1, we provide one example to demonstrate the visual distortion from these two aspects. From this figure, we can see that the SR image is smoother compared with the LR image, which

(a) SR image (b) HR image

Fig. 1. LR and SR image samples: the SR image is obtained from the study [24]. The energy change and texture variation are computed based on the differences between corresponding image patches in LR and SR images.

can be reflected by the energy change of low-frequency regions in the SR image. It is also the global distortion in the SR image. Furthermore, from the small patch in Fig. 1(b), we can observe that there is much visual distortion in high-frequency regions along the eye, nose, *etc.* Compared with the LR image, we can use the texture variation to represent the local information change in high-frequency regions, which can be also considered as the local distortion. Thus, we propose to measure the visual distortion of HR images from these two aspects: the global visual information degradation from energy change, and the local visual distortion from texture variation.

The framework of the proposed method is shown as Fig. 2. From this figure, we can see that the proposed method first obtains the pixel correspondence between LR and HR images. Then the energy change and texture variation are estimated by the feature difference between LR and HR images. The final RRIQA-SR score is predicted by considering both the energy change and texture variation in HR images. We will explain the framework of the proposed method step by step as follows.

3.2. Pixel correspondence

Since the size of the HR image becomes larger due to generated image pixels from the LR image, the pixel correspondence between LR and HR images is missing. In general, we do not know the algorithm for super-resolution construction. The difference between the LR and HR images from image super-resolution is mainly caused by structure change due to generated image pixels in HR images. Assume the reference and distorted images aligned well, traditional perceptual metrics such as SSIM mainly calculate the pixel-to-pixel difference for visual quality assessment. The classical distortions such as Gaussian noise, compression artifact, and contrast change can be regarded as the intensity changes, which can be captured by the direct subtraction of the reference and distorted images. However, it is impossible to predict the visual quality of HR image from image super-resolution by direct subtraction of LR and HR images, since their resolutions are different. Thus, we have to obtain the pixel correspondence first before the similarity calculation. In this study, the pixel correspondence between LR and HR images is modeled by the Markov Random Field (MRF) [31] in energy minimization framework. The SIFT descriptors [32] has been proved to be robust for pixel matching across different scenes. Here, we use the SIFT descriptor as the features for pixel correspondence prediction.

3.3. Energy change and texture variation

After pixel correspondence, we calculate the energy change and texture variation between image patches in LR and HR images. Given a LR image I_{LR} and its corresponding HR image I_{HR}, their sizes are denoted as $M_{LR} \times N_{LR}$ and $M_{HR} \times N_{HR}$. Thus, the resizing factor α can be calculated as: $\alpha = M_{HR}/M_{LR}$. The computation of energy change and texture variation between LR and HR images are given as follows.

$$F_k(I_{LR}, I_{HR}) = \sum_{(b,b')} f_k(b, b') \tag{1}$$

where $k \in \{1, 2\}$ represents the energy or texture feature; f_k denotes the function to compute energy change or texture variation. b and b' are the corresponding image patches centering at the pixel pair p and p' in LR and HR images, respectively. Please note that the size of image patch b' is α times of that of b, as shown in the small patches denoted by white squares in Fig. 1. Here, for each image pixel p in the LR image, we extract one image patch pair based on pixel correspondence for the energy change and texture variation calculation in Eq. (1). The overall perceptual similarity between the LR and HR images is represented by the sum of similarities of all patch pairs in LR and HR images.

During the past decades, Discrete Cosine Transform (DCT) has been widely used for feature representation in various image processing applications [22,23]. It is well known that the DC coefficient includes most of the image energy and represents the energy of the image, while AC coefficients represent the frequency components in images [23]. Here, we use the DC coefficient to represent the energy feature of each image patch, while the texture feature is extracted from AC coefficients.

3.3.1. Energy change estimation

Given any image patch pair b and b' from the LR and HR images, we first calculate their DC coefficients by DCT as D and D' for image patches b and b', respectively. The average energy change between this image patch pair can be computed as:

$$f_e(b, b') = \frac{2m_D m_{D'} + C_1}{m_D^2 + m_{D'}^2 + C_1} \tag{2}$$

where C_1 is a constant; m_D and $m_{D'}$ represent the average energy values in image patches b and b', respectively, and they are computed as:

$$m_D = \frac{D}{N^2} \tag{3}$$

$$m_{D'} = \frac{D'}{N'^2} \tag{4}$$

where $N \times N$ and $N' \times N'$ denote the sizes of image patches i and j, respectively.

In Fig. 3, we provide the similarity map from energy change in the fourth column. From this map, we can see that the energy change capture the information degradation globally, which

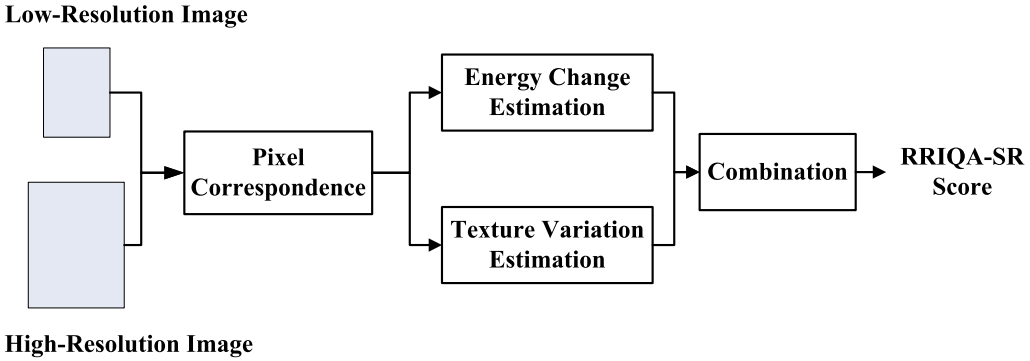


Fig. 2. The framework of the proposed method.

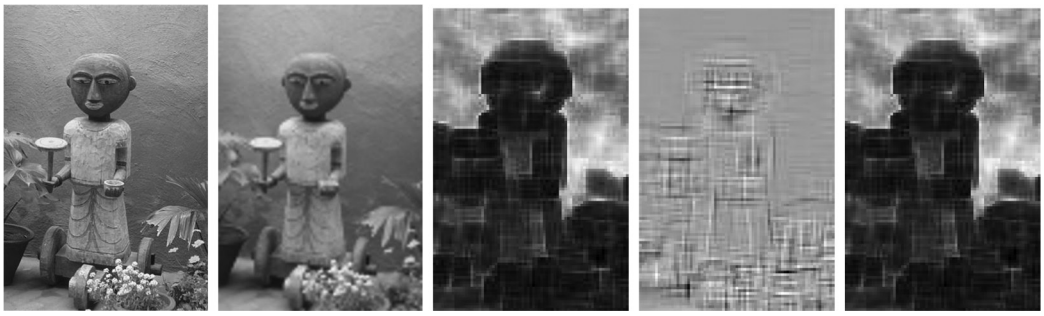


Fig. 3. The visual samples of similarity maps from different components in the proposed method. First column: the ground truth image; second column: SR image generated from the LR image; third column: similarity map from texture variation; fourth column: similarity map from energy change; fifth column: overall similarity map from the proposed method. Please note that the contrast of the similarity maps is enlarged for better visual experiences.

demonstrates that it can represent the visual distortion of low-frequency regions in SR images.

3.3.2. Texture variation estimation

From Eq. (2), we can calculate the average energy change between image patch pair in LR and HR images. For the texture variation between image patch pairs in LR and HR images, we use AC coefficients to represent the texture feature. For any image patch b with size $N_b \times N_b$ in the LR image, it has $N_b^2 - 1$ AC coefficients: $A = \left\{A_1, A_2, A_3, \ldots, A_{N_b^2-1}\right\}$. For any image patch $N_{b'}$ in the HR image, there are $N_{b'}^2 - 1$ AC coefficients: $A' = \left\{A'_1, A'_2, A'_3, \ldots, A'_{N_{b'}^2-1}\right\}$. The texture variation between image patches in LR and HR images can be calculated by the differences of the mean and standard deviation values of AC coefficients. The texture variation by patch differences between image patches b and b' can be computed as follows.

$$f_t(b,b') = \frac{(2m_A m_{A'} + C_2)(2d_A d_{A'} + C_3)}{\left(m_A^2 + m_{A'}^2 + C_2\right)\left(d_A^2 + d_{A'}^2 + C_3\right)} \tag{5}$$

where m_A and $m_{A'}$ are the mean values of the vectors A and A', respectively; d_A and $d_{A'}$ denote the standard deviation of the vectors A and A', respectively; C_2 and C_3 are constant values. The mean and standard deviation values are computed as follows.

$$m_A = \frac{\sum_{p=1}^{N_b^2-1} A_p}{N_b^2 - 1} \tag{6}$$

$$m_{A'} = \frac{\sum_{p=1}^{N_{b'}^2-1} A'_p}{N_{b'}^2 - 1} \tag{7}$$

$$d_A = \sqrt{\frac{1}{N_b^2 - 1}\sum_{p=1}^{N_b^2-1}\left(A_p - m_A\right)^2} \tag{8}$$

$$d_{A'} = \sqrt{\frac{1}{N_{b'}^2 - 1}\sum_{p=1}^{N_{b'}^2-1}\left(A'_p - m_{A'}\right)^2} \tag{9}$$

In Fig. 3, we provide the similarity map from texture variation in the third column. From this figure, we can see that the local distortion is well captured by texture variation, especially for the regions with complex texture (high-frequency regions). Thus, we can use the energy change and texture variation of the HR image to estimate the visual distortion of SR images globally and locally according to Eqs. (2) and (5), respectively. In the next subsection, we will introduce how to predict the visual quality of HR images based on these two components.

3.4. Overall quality prediction

As indicated previously, the energy change in HR images would cause the global visual information degradation to the image, while the texture variation would bring into local distortion to high-frequency regions. Thus, we predict the visual quality of HR images by combining these two components as follows.

$$Q = F_e * F_t \tag{10}$$

where F_e and F_t represent the pooling values of estimated energy change and texture variation from all patch pairs between LR and HR images.

In Fig. 3, the overall similarity map from energy change and texture variation is given in the last column. From these similarity maps, we can see that the overall similarity map is very similar with the similarity map from texture variation, which demonstrate that the energy change from different local patches are almost similar in SR images. More analysis is provided in the experiment section.

4. Experimental results

In this section, we provide the experimental results for the performance evaluation of the proposed RRIQA-SR. First, we give the evaluation methodology of the comparison experiments, including the used databases and evaluation methods. Then we analyze the influence of each component in the proposed method for quality evaluation of SR images. Following this, the existing IQA metrics are used to conduct the comparison experiments for the performance evaluation of the proposed method.

4.1. Evaluation methodology

We use the database with subjective scores in [8] to do the comparison experiment. Although the ground truth HR images are available in [8], we have only used the generated LR images from them, not these ground truth images, for the proposed RRIQA-SR; these ground truth HR images are also used for performance evaluation for the existing full-references IQA metrics under comparison. These ground truth HR images covering a wide range of high-frequency levels are selected from Berkeley segmentation dataset [25], where the images are with diverse content obtained in a professional photographic style. The ground truth images are first used to generate LR images for image super-resolution.

There are ten ground truth images used in the subjective test. For each ground truth image, nine LR images are created under three scaling factors and three Gaussian kernel widths. The SR images are generated from LR images by six existing single frame

super-resolution algorithms. Thus, there are 540 SR images in total in this database. Thirty participants were involved in the subjective test, evaluating the 540 SR images without knowing the ground truth images or image super-resolution methods. During the subjective test, SR images were displayed randomly to avoid the bias to favor specific methods and participants were asked to give a perceptual score between 0 and 10 for each SR image [8]. The subjective perceptual quality of SR images is represented by the MOS, which is the average of the subjective scores over 30 participants.

The performance of the proposed method can be evaluated by the correlation between subjective and objective scores. In this study, we use three common methods to calculate the correlation between the subjective and objective scores: Pearson linear correlation coefficient (PLCC), Spearman rank-order correlation coefficient (SRCC) and Kendall rank correlation coefficient (KRCC). PLCC is computed as the correlation between subjective and objective scores with a nonlinear mapping. Given the ith image in the database with size N, its subjective and objective scores are s_i and o_i, respectively. We use a nonlinear function to o_i as follows [52]:

$$o_i' = \alpha_1\left[\frac{1}{2} - \frac{1}{1+exp(\alpha_2(o_i-\alpha_3))}\right] + \alpha_4 o_i + \alpha_5 \tag{11}$$

where α_1 to α_5 are parameters found numerically with a nonlinear regression to maximize the correlations between subjective and objective scores. PLCC can be estimated as:

$$PLCC = \frac{\sum_i (o_i' - \overline{o'})(s_i - \overline{s})}{\sqrt{\sum_i (o_i' - \overline{o'})\sum_i (s_i - \overline{s})}} \tag{12}$$

where $\overline{o'}$ is the mean value of o_i'; $\overline{s}$ is the mean value of s_i.

SRCC and KRCC can be computed as follows:

$$SRCC = 1 - \frac{6\sum_{i=1}^{N} e_i^2}{N\left(N^2-1\right)} \tag{13}$$

where e_i is the difference between the ith image's ranks in subjective and objective results.

$$KRCC = \frac{N_c - N_d}{\frac{1}{2}N(N-1)} \tag{14}$$

where N_c and N_d are the numbers of concordant and discordant pairs in the dataset, respectively.

PLCC, SRCC and KRCC can be computed as Eqs. (12)–(14). PLCC can be used to evaluate the quality prediction accuracy, while SRCC an KRCC can be used to measure the monotonicity of quality prediction [53]. Generally, a better IQA metric can obtain higher PLCC, KRCC and SRCC values.

4.2. Experiment 1: influence of each component in the proposed method

From Eq. (10), we can see that there are two components in the proposed method: the energy change F_e, and the texture variation F_t. Here, we first conduct the comparison experiment to evaluate the influence of these two components. In this subsection, we use energy change, texture variation, and both of them to compute the objective scores of HR images. Then the objective scores from these three methods are used to be compared with subjective scores. Experimental results of PLCC, KRCC and SRCC are listed in Table 1.

Table 1
Performance evaluation of the proposed method on two components.

Components	Energy Change	Texture Variation	Proposed
PLCC	0.5803	0.6880	0.8052
KRCC	0.4092	0.4996	0.5885
SRCC	0.6001	0.6958	0.8035

From Table 1, we can see that the quality evaluation results from the component texture variation F_t can obtain higher correlation with subjective data than those from the component energy change F_e. This demonstrates that the texture variation would influence the overall visual quality of HR images more than the energy change. This is reasonable, since the human visual system is always much sensitive to high-frequency regions. And thus, the visual distortion in high-frequency regions such as edges is more obvious than the overall information degradation in HR images. As shown in Table 1, the proposed method by combining these two components can obtain much better performance than each component energy change or texture variation.

Here, we also provide some visual samples of similarity maps in Fig. 4. In this figure, we provide the similarity maps from texture variation, energy change and the proposed method including both these two components. From the first row of this figure, we can see that the similarity map from texture variation mainly capture the visual distortion from the texture changes in regions, especially the regions with complex texture. For the energy change component, its similarity map mainly represents the overall information degradation in SR images. By comparing the visual samples from the first and second rows, we can see that the visual distortion in similarity map from texture variation in the second row is less than that in the first row. This can be also observed from the SR images in the first and second rows. By comparing the SR images in the first and second rows, we can see that the visual distortion in the first row is obviously larger than that in the second row, especially for the high-frequency regions. Thus, the texture variation captures the local visual distortion in high-frequency regions of SR images well.

From Fig. 4, we can also find that the overall similarity map is similar with that from texture variation for each SR image. This demonstrates that the energy change of different local patches in SR images are almost the same. This can be confirmed by the similarity maps from energy change, which show that energy change of most regions are the same. Thus, the defined energy change of the proposed method can be used for capturing the visual distortion of low-frequency regions, which is a measure of global visual distortion in the SR image. In contrast, the texture variation of the proposed method can be used to represent the visual distortion of high-frequency regions, which is a measure of local visual distortion in the SR image.

4.3. Experiment 2: comparison with existing related metrics

To further demonstrate the performance of the proposed method, we have conducted the comparison experiments by using some existing IQA metrics. Here, we use the NSS-SR metric [20] designed specifically for image super-resolution in the comparison experiments. Please note that metric also use the LR images as the reference information and thus it is also a RR metrics for image super-resolution. The following popular full reference quality metrics are also used in performance evaluation due to the available ground truth information in the database: PSNR, SSIM [18], multi-scale SSIM (MSSSIM) [26], noise quality measure (NQM) [28], visual information fidelity (VIF) [29], and the most apparent distortion (MAD) [30]. We obtained the available source code of these existing studies from the corresponding authors. Experimental results of PLCC, SRCC and KRCC are shown in Table 2.

From Table 2, we can see that SSIM can obtain better performance than PSNR, similar with visual quality for general images [18]. This demonstrates that perceptual consideration of structure

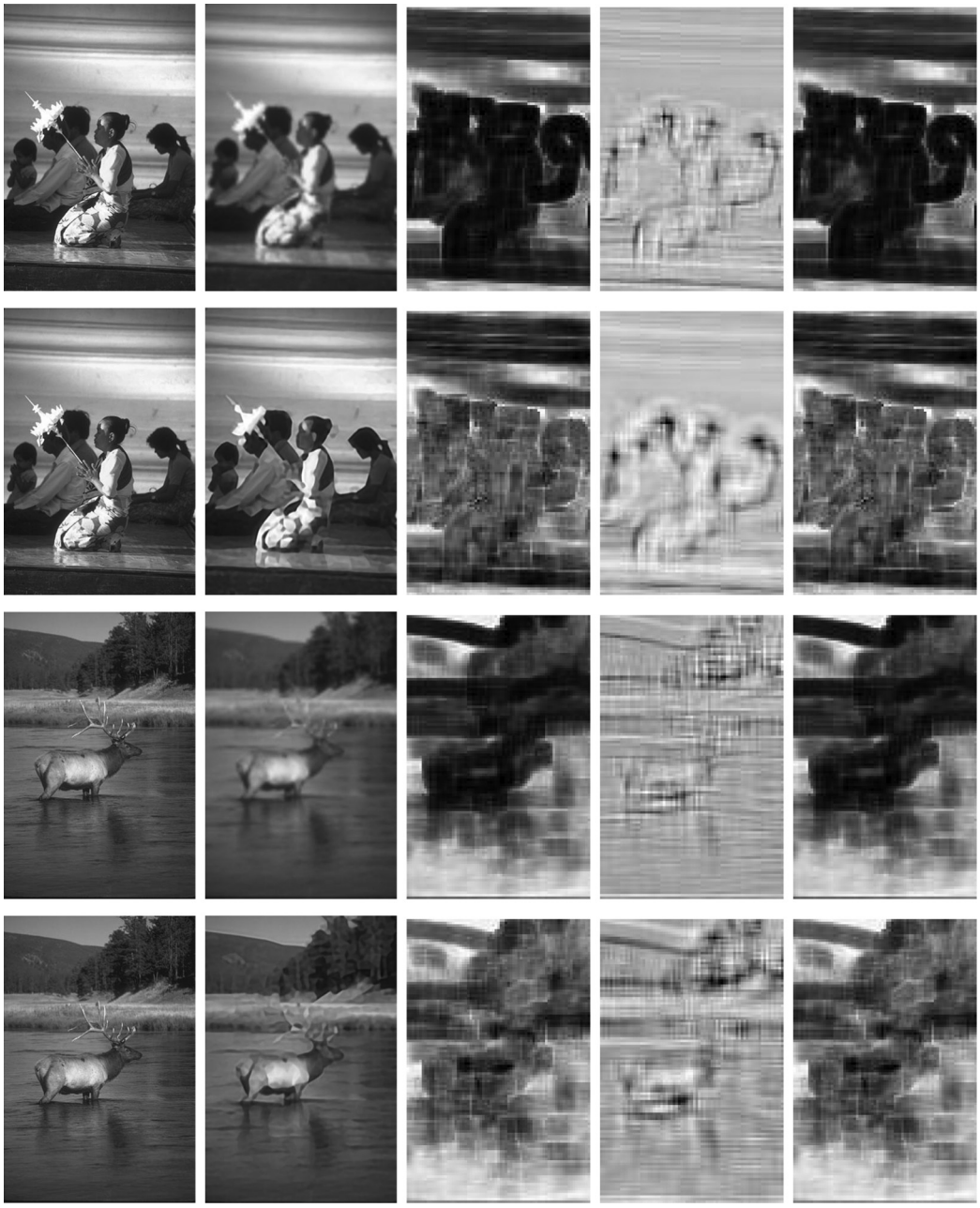
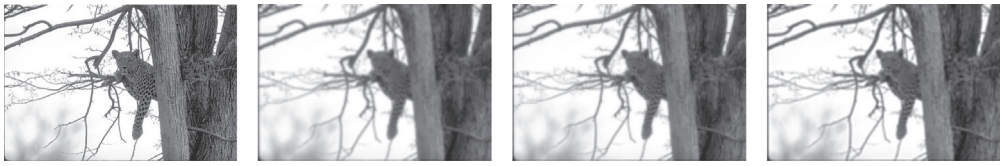


Fig. 4. The visual samples of similarity maps from two components in the proposed method. First column: the ground truth images; second column: SR images generated from the corresponding LR images; third column: similarity maps from texture variation; fourth column: similarity maps from energy variation; fifth column: overall similarity maps from the proposed method. Please note that the contrast of the similarity maps is enlarged for better visual experiences.

Table 2
Performance evaluation of the proposed method.

	PSNR	SSIM	MSSSIM	NQM	VIF	MAD	NSS-SR	Proposed
PLCC	0.5145	0.6702	0.7504	0.7940	0.5351	0.7924	0.1614	0.8053
KRCC	0.3296	0.4502	0.5325	0.5703	0.2786	0.5523	0.0917	0.5885
SRCC	0.4760	0.6203	0.7096	0.7632	0.5226	0.7363	0.1343	0.8035

information is useful in quality prediction of image super-resolution. Compared with SSIM and PSNR, MSSSIM can obtain better performance in quality prediction of HR images. The reason is that MSSSIM uses more high-frequency information through the multi-scale implementation for quality predication. As demonstrated previously, human perception is more sensitive to the visual distortion in high-frequency regions than that in low-frequency regions. NQM and MAD can obtain better performance in quality prediction of HR images than VIF and MSSIM. In both NQM and MAD, the contrast sensitivity and contrast masking are used to model the human visual perception in different frequencies. Thus, visual distortion in high-frequency regions of HR images



(a) Ground Truth (b) SR image 1 (c) SR image 2 (d) SR image 3

Fig. 5. The ground truth image (a) and SR images. (b) MOS: 1.8462, RRIQA-SR:0.4397, MSSSIM: 0.8414, NQM: 17.6, PSNR: 19.3077; (c) MOS: 2.0769, RRIQA-SR: 0.4831, MSSSIM: 0.8737, NQM: 22.3425, PSNR: 19.9234; (d) MOS: 2.1538, RRIQA-SR:0.4853, MSSSIM: 0.8631, NQM: 18.7231, PSNR: 19.7341.

(a) Ground Truth (b) SR image 1 (c) SR image 2 (d) SR image 3

Fig. 6. The ground truth image (a) and SR images. (b) MOS: 0.889, RRIQA-SR:0.2690, MSSSIM: 0.8657, NQM: 17.15, PSNR: 21.18; (c) MOS: 1, RRIQA-SR: 0.2764, MSSSIM: 0.8660, NQM: 17.1, PSNR: 21.1833; (d) MOS: 1.2222, RRIQA-SR:0.3596, MSSSIM: 0.8488, NQM: 15.4881, PSNR: 20.4188.

can be well measured by NQM and MAD. From Table 2, we can see that NSS-SR obtains the lowest performance among the compared quality metrics. Although NSS-SR is designed for image super-resolution, the NSS models used in that metric are built more specifically for image interpolation [20]. The used SR images are created by using various image super-resolution algorithms rather than image interpolation [8]. Thus, the NSS-SR cannot work well in this database. Compared with other existing studies, we can see that the proposed RRIQA-SR can get higher PLCC, KRCC, and SRCC values than other compared metrics. This demonstrates that the proposed method can obtain better performance in visual quality prediction for image super-resolution than other metrics, even for FR metrics. The reason is that the proposed method can measure the visual distortion of low-frequency region and high-frequency region by energy change and texture variation, respectively.

In Fig. 5, we provide some SR image samples with subjective and objective scores calculated from different IQA metrics. From this figure, we can see that all used quality metrics can predict the consistent quality of SR images 1 and 2 with subjective data. However, for SR image 3 with better quality than SR image 2, the metrics of MSSSIM, NQM and PSNR cannot predict the visual quality accurately. In contrast, the proposed RRIQA-SR can predict the visual quality of all these images consistently with the subjective data. Fig. 6 also provides some visual samples with subjective and objective scores from different metrics. From this figure, we can also see that the proposed method can predict the visual quality of HR images more consistently than other existing metrics.

5. Conclusion

In this paper, a novel RRIQA-SR has been built for image super-resolution, since reduced-reference IQA is the most meaningful and practical IQA for this application. MRF model is first used to estimate pixel correspondence between LR and HR images. Then the energy change and texture variation from image patch pairs between LR and HR images are calculated to predict visual distortion of low-frequency and high-frequency regions in HR images, respectively. The visual quality of HR images is predicted by considering both energy change and texture variation, which can capture the global and local distortion in HR images, respectively. Experimental results show that the proposed RRIQA-SR method can obtain better performance than other quality metrics, even some FR quality metrics. In the future, we will investigate how to use the proposed RRIQA-SR to optimize image super-resolution algorithms.

Acknowledgement

This work was supported in part by the Natural Science Foundation of China under Grant 61571212 and 61822109, the Natural Science Foundation of Jiangxi Province under Grant 20181BBH80002, and the Fok Ying-Tong Education Foundation of China under Grant 161061.

References

[1] K. Nasrollahi, T.B. Moeslund, Super-resolution: a comprehensive survey, Mach. Vis. Appl. 25 (6) (2014) 1423–1468.
[2] J. Tian, K.-K. Ma, A survey on super-resolution imaging, SIViP 5 (3) (2011) 329–342.
[3] M. Shen, C. Wang, P. Xue, W. Lin, Performance of reconstruction-based super-resolution with regularization, J. Visual Commun. Image Represent. 21 (7) (2010) 640–650.
[4] C. Wang, P. Xue, W. Lin, Improved super-resolution reconstruction from video, IEEE Trans. Circuits Syst. Video Technol. 16 (11) (2006) 1411–1422.
[5] R.G. Keys, Cubic convolution interpolation for digital image processing, IEEE Trans. Acoust. Speech Signal Process. 29 (6) (1981) 1153–1160.
[6] R. Fattal, Image upsampling via imposed edge statistics, ACM Trans. Graph. 26 (3) (2007) 95.

[7] J. Sun, Z. Xu, H.Y. Shum, Image super-resolution using gradient profile prior, in: IEEE International Conference on Computer Vision and Pattern Recognition, 2008.
[8] C.-Y. Yang, C. Ma, M.-H. Yang, Single-image super-resolution: a benchmark, in: European Conference on Computer Vision, 2014.
[9] Y. Zhang, J. Liu, W. Yang, Z. Guo, Image super-resolution based on structure-modulated sparse representation, IEEE Trans. Image Process. 24 (9) (2015) 2797–2810.
[10] W. Dong, L. Zhang, G. Shi, X. Wu, Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization, IEEE Trans. Image Process. 20 (7) (2011) 1838–1857.
[11] Y.-Q. Zhang, Y. Ding, J. Liu, Z. Guo, Guided image filtering using signal subspace projection, IET Image Proc. 7 (3) (2013) 270–279.
[12] K. Ni, T. Nguyen, Image superresolution using support vector regression, IEEE Trans. Image Process. 16 (6) (2007) 1596–1610.
[13] J. Ren, J. Liu, Z. Guo, Context-aware sparse decomposition for image denoising and super-resolution, IEEE Trans. Image Process. 22 (4) (2013) 1456–1469.
[14] J. Yang, J. Wright, T.S. Huang, Y. Ma, Image super-resolution via sparse representation, IEEE Trans. Image Process. 19 (11) (2010) 2861–2873.
[15] C. Dong, C.C. Loy, K. He, X. Tang, Learning a deep convolutional network for image super-resolution, in: European Conference on Computer Vision, 2014.
[16] Z. Wang, A.C. Bovik, Model image quality assessment, in: Syntheses Lectures on Image, Video and Multimedia Processing, Morgan and Claypool Publishers, 2006.
[17] W. Lin, C.-C. Jay Kuo, Perceptual visual quality metrics: a survey, J. Vis. Commun. Image Represent. 22 (4) (2011) 297–312.
[18] 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.
[19] J. Wu, W. Lin, G. Shi, A. Liu, Perceptual quality metric with internal generative mechanism, IEEE Trans. Image Process. 22 (1) (2013) 43–54.
[20] H. Yeganeh, M. Rostami, Z. Wang, Objective quality assessment for image super-resolution: a natural scene statistics approach, in: IEEE International Conference on Image Processing, 2012.
[21] A.R. Reibman, R.M. Bell, S. Gray, Quality assessment for super-resolution image enhancement, in: IEEE International Conference on Image Processing, 2006.
[22] N. Ahmed, T. Natarajan, K.R. Rao, Discrete cosine transform, IEEE Trans. Comput. 23 (1) (1974) 90–93.
[23] R.C. Gonzalez, R.E. Woods, Digital Image Processing, third ed., Prentice Hall, 2008.
[24] C.Y. Yang, M.H. Yang, Fast direct super-resolution by simple functions, in: IEEE International Conference on Computer Vision, 2013.
[25] D. Martin, C. Fowlkes, D. Tal, J. Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, in: IEEE International Conference on Computer Vision, 2001.
[26] Z. Wang, E. Simoncelli, A.C. Bovik, Multi-scale structural similarity for image quality assessment, in: IEEE Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems, and Computers, 2003.
[28] N. Damera-Venkata, T.D. Kite, W.S. Geisler, B.L. Evans, A.C. Bovik, Image quality assessment based on a degradation model, IEEE Trans. Image Process. 9 (4) (2000) 636–650.
[29] H.R. Sheikh, A.C. Bovik, Image information and visual quality, IEEE Trans. Image Process. 15 (2) (2006) 430–444.
[30] 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.
[31] S.Z. Li, Markov Random Field Modeling in Image Analysis, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2001.
[32] D.G. Lowe, Object recognition from local scale-invariant features, in: IEEE International Conference on Computer Vision, 1999.
[33] L.C. Pickup, D.P. Capel, S.J. Roberts, A. Zisserman, Bayesian methods for image super-resolution, Comput. J. 52 (1) (2009) 101–113.
[34] L.J. Karam, N.G. Sadaka, R. Ferzli, Z.A. Ivanovski, An efficient selective perceptual-based super-resolution estimator, IEEE Trans. Image Process. 20 (12) (2011) 3470–3482.
[35] X. Zhang, M. Tang, R. Tong, Robust super resolution of compressed video, The Visual Computer 28 (12) (2012) 1167–1180.
[36] Z. Wei, K.-K. Ma, Contrast-guided image interpolation, IEEE Trans. Image Process. 22 (11) (2013) 4271–4285.
[37] J. Mairal, F. Bach, J. Ponce, G. Sapiro, A. Zisserman, Non-local sparse models for image restoration, in: IEEE International Conference on Computer Vision, 2009.
[38] K.I. Kim, Y. Kwon, Single-image super-resolution using sparse regression and natural image prior, IEEE Trans. Pattern Anal. Mach. Intell. 32 (6) (2010) 1127–1133.
[39] X. Li, M.T. Orchard, 'New edge-directed interpolation', IEEE Trans. Image Process. 10 (10) (2001) 1521–1527.
[40] L. Zhang, X. Wu, An edge-guided image interpolation algorithm via directional filtering and data fusion, IEEE Trans. Image Process. 15 (8) (2006) 2226–2238.
[41] L.-W. Kang, C.-C. Hsu, B. Zhang, C.-W. Lin, Learning-based joint super-resolution and deblocking for a highly compressed image, IEEE Trans. Multimedia 17 (7) (2015) 921–934.
[42] Y.-Q. Zhang, Y. Ding, J.-S. Xiao, J. Liu, Z. Guo, Visibility enhancement using an image filtering approach, EURASIP J. Adv. Signal Process. 2012 (220) (2012) 1–6.
[43] Y. Fang, W. Lin, S. Winkler, Review of existing QoE methodologies, in: C.W. Chen, P. Chatzimisios, T. Dagiuklas, L. Atzori (Eds.), Chapter 3 in Multimedia Quality of Experience (QoE): Current Status and Future Requirements, Wiley, 2015.
[44] A. Liu, W. Lin, M. Narwaria, Image quality assessment based on gradient similarity, IEEE Trans. Image Process. 21 (4) (2012) 1500–1512.
[45] 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.
[46] D.M. Chandler, S.S. Hemami, VSNR: a wavelet-based visual signal-to-noise ratio for natural images, IEEE Trans. Image Process. 16 (9) (2007) 2284–2298.
[47] Q. Li, Z. Wang, Reduced-reference image quality assessment using divisive normalization-based image representation, IEEE J. Sel. Top. Signal Process. 3 (2) (2009) 202–211.
[48] J. Wu, W. Lin, G. Shi, A. Liu, Reduced-reference image quality assessment with visual information fidelity, IEEE Trans. Multimedia 15 (7) (2013) 1700–1705.
[49] S. Wang, X. Zhang, S. Ma, W. Gao, Reduced reference image quality assessment using entropy of primitives, in: The 30th Picture Coding Symposium, 2013, pp. 193–196.
[50] G. Zhai, X. Wu, X. Yang, W. Lin, W. Zhang, A psychovisual quality metric in free-energy principle, IEEE Trans. Image Process. 21 (1) (2012) 41–52.
[51] Y. Fang, K. Ma, Z. Wang, W. Lin, Z. Fang, G. Zhai, No-reference quality assessment of contrast-distorted images based on natural scene statistics, IEEE Signal Process. Lett. 22 (7) (2015) 838–842.
[52] Z. Wang, Q. Li, Information content weighting for perceptual image quality assessment, IEEE Trans. Image Process. 20 (5) (2011) 1185–1198.
[53] VQEG, Final report from the video quality experts group on the validation of objective models of video quality assessment, Apr. 2000, available at [Online]. Available: <http://www.vqeg.org/>.
[54] Y. Zhang, Y. Zhang, J. Zhang, Q. Dai, CCR: clustering and collaborative representation for fast single image super-resolution, IEEE Trans. Multimedia 18 (3) (2015) 405–417.
[55] Y. Fang, J. Yan, J. Liu, S. Wang, Q. Li, Z. Guo, Objective quality assessment of screen content images by uncertainty weighting, IEEE Trans. Image Process. 26 (4) (2017) 2016–2027.
[56] L. Xing, L. Cai, H. Zeng, J. Chen, J. Zhu, J. Hou, A multi-scale contrast-based image quality assessment model for multi-exposure image fusion, Signal Process. 145 (2018) 233–240.
[57] Z. Ni, L. Ma, H. Zeng, J. Chen, C. Cai, K.-K. Ma, ESIM: edge similarity for screen content image quality assessment, IEEE Trans. Image Process. 26 (10) (2017) 4818–4831.
[58] Y. Fang, J. Yan, L. Li, J. Wu, W. Lin, No reference quality assessment for screen content images with both local and global feature representation, IEEE Trans. Image Process. 27 (4) (2018) 1600–1610.
[59] Y. Fang, J. Liu, Y. Zhang, W. Lin, Z. Guo, Quality assessment for image super-resolution based on energy change and texture variation, in: IEEE International Conference on Image Processing, 2016.