

QUALITY ASSESSMENT FOR IMAGE SUPER-RESOLUTION BASED ON ENERGY CHANGE AND TEXTURE VARIATION

Yuming Fang^{1,2}, Jiaying Liu^{2,*}, Yabin Zhang³, Weisi Lin³, and Zongming Guo²

¹School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, China

²Institute of Computer Science and Technology, Peking University, Beijing, China

³School of Computer Engineering, Nanyang Technological University, Singapore

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. First, we use the Markov Random Field (MRF) to model the pixel correspondence between LR and high-resolution (HR) images. Based on the pixel correspondence, we predict the perceptual similarity between image patches of LR and HR images by two components: the energy change and texture variation. The overall quality of HR images is estimated by the perceptual similarity between local image patches of LR and HR images. Experimental results demonstrate that the proposed method can obtain better performance of quality prediction for HR images than other existing ones, even including some full-reference (FR) metrics.

Index Terms— Image quality assessment (IQA), image super-resolution, reduced-reference (RR) quality assessment, energy change, texture variation

1. INTRODUCTION

The image super-resolution technique aims to construct a high-resolution (HR) image with one or several given low-resolution (LR) images. It has been widely used in various applications, including medical image processing, infrared imaging, face/iris recognition, image editing, *etc.*. During the past decades, there are numerous image super-resolution algorithms proposed [1]- [10]. According to the number of available LR images, the super-resolution algorithms can be classified into two categories: multi-frame super-resolution and single-frame super-resolution approaches [6].

For single-image super-resolution methods, there have been many different approaches proposed previously. Traditional interpolation based methods try to reconstruct the HR image by a base function, including bilinear, bicubic and nearest neighbor algorithms [3]. Generally, these approaches

are simple and efficient, but there are serious aliasing artifacts and blurring distortions along edges and high-frequency regions due to the pixel interpolation operation.

To overcome the drawback from interpolation based methods, many advanced image super-resolution algorithms have been proposed. The reconstruction-based methods generate HR images by a regularized cost function with certain prior knowledge. The prior information used in these methods includes the edge information [4], sparsity priors [7], *etc.* The example learning-based methods reconstruct 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 [5]. There have been various learning techniques used for mapping functions, including support vector regression [8], sparse dictionary representation [6, 9], deep learning [10], *etc.*

For these existing studies introduced above, the performance of super-resolution algorithms is validated by small-scale subjective tests. Currently, much less has been done to assess the visual quality of HR images quantitatively. As we know, the subjective test is expensive, time-consuming, and cannot be embedded into super-resolution algorithms for optimization purpose. Thus, objective quality assessment is much desired for quality evaluation of HR images.

There has been much progress in the area of image quality assessment (IQA) [11, 25]. However, traditional IQA metrics such as peak signal-to-noise-ratio (PSNR), the structural similarity (SSIM) [12] and the internal generative mechanism (IGM) [13], cannot be used in super-resolution applications, since they need the sizes of the reference and distorted images to be the same. Currently, there are only a few IQA studies investigating the visual quality assessment of HR images [5, 14, 15]. The authors in [14] proposed an objective IQA metric based on natural scene statistics (NSS). However, that NSS based method is mainly designed for interpolated natural images [14, 16]. 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 [5]. The full-reference IQA metrics such

*Corresponding Author. This work was supported by the NSFC (No. 61571212), 863 Program of China (No. 2014AA015205), NSF of Beijing (No.4142021) and NSF of Jiangxi Province (No. 20151BDH80003).

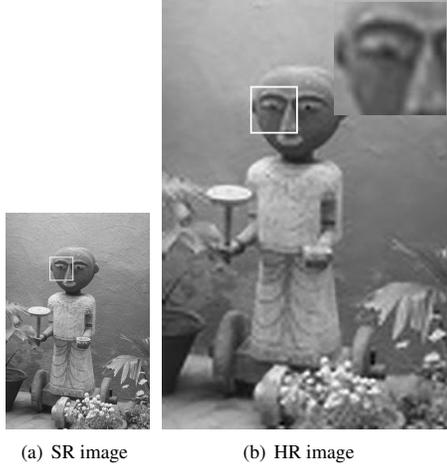


Fig. 1. SR and HR image samples: the HR image is obtained from the study [19]. The energy change and texture variation are computed based on the differences between corresponding image patches in LR and HR images.

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 not 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.

In this study, we propose a novel reduced-reference (RR) quality metric for image super-resolution (RRIQA-SR); 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. We focus on the visual quality prediction of single-frame super-resolution. For image super-resolution, the global structural information in the generated HR image should be based upon that in the LR image. To measure the visual quality in HR images, the proposed method first uses a Markov Random Field (MRF) to model the pixel correspondence. Then the energy and texture features are extracted from image patches to compute the perceptual similarity between LR and HR images, which is further adopted to predict the overall quality of HR images. Experimental results show that the proposed RRIQA-SR can obtain better performance in quality prediction of HR images than other existing ones.

2. PROPOSED METHOD

2.1. Overview

In image super-resolution construction, the visual information of the generated HR image should be highly similar with that of the original LR image. For the reconstructed HR im-

ages, the visual distortion brought into during image super-resolution operation is mainly caused from the two aspects: one is the overall energy change from a LR image to its generated HR image, while the other is the information degradation in high-frequency regions such as edges, corners, *etc.*

In Fig. 1, we provide one example to demonstrate the visual distortion from these two aspects. From this figure, we can see that the HR image is smoother compared with the LR image, which can be reflected by the energy change during image super-resolution operation. 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 information change in the high-frequency regions. Thus, we propose to measure the visual distortion of HR images from these two aspects: the visual information degradation from energy change, and the visual distortion from texture variation.

2.2. 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, since in general, we do not know the algorithm for super-resolution construction. To calculate the local distortion in HR images, we model the pixel correspondence between LR and HR images with the Markov Random Field (MRF) [27] in energy minimization framework as follows [28]:

$$E = \sum_p \min d(g(p), g(p')) + \omega \sum_{(p,q') \in \Phi} \min(|\mu(p) - \mu(q)| + |\nu(p) - \nu(q)|) \quad (1)$$

where $d(g(p), g(p'))$ represents the distance between the features at pixel pair p and p' ; $(\mu(p), \nu(p))$ is the flow vector at pixel p ; Φ is the set containing the spatial neighbors centering at pixel p ; ω is a parameter to determine the relative importance. In Eq. (1), the first term is used to obtain the pixel pair with minimum feature change, while the second term is adopted to guarantee the smoothness of the pixel correspondence. The scale-invariant feature transform (SIFT) descriptors [29] has been proved to be robust for pixel matching across different scenes. Here, we use the SIFT descriptor as the features $g(p)$ and $g(p')$ for pixel pair p and p' .

2.3. 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.

$$S_k(I_{LR}, I_{HR}) = \sum_{(b,b')} f_k(b, b') \quad (2)$$

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. (2). 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 [17, 18]. 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 [18]. Here, we use the DC coefficient to represent the energy feature of each image patch, while the texture feature is extracted from AC coefficients.

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:

$$f_e(b, b') = \frac{2m_D m_{D'} + C_1}{m_D^2 + m_{D'}^2 + C_1} \quad (3)$$

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

From Eq. (3), 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: $A = \{A_1, A_2, A_3, \dots, A_{N_b^2-1}\}$. For any image patch $N_{b'}$ in the HR image, there are $N_{b'}^2 - 1$ AC coefficients: $A' = \{A'_1, A'_2, A'_3, \dots, A'_{N_{b'}^2-1}\}$. 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 the patch differences between image patches b and b' can be computed as follows.

$$f_t(b, b') = \frac{(2m_A m_{A'} + C_2)(2d_A d_{A'} + C_3)}{(m_A^2 + m_{A'}^2 + C_2)(d_A^2 + d_{A'}^2 + C_3)} \quad (4)$$

where m_A and $m_{A'}$ are the mean values of the vectors A and

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

Components	Energy Change	Texture Variation	Proposed
KRCC	0.4092	0.4996	0.5885
SRCC	0.6001	0.6958	0.8035

A' , respectively; d_A and $d_{A'}$ denote the standard deviation of the vectors A and A' , respectively; C_2 and C_3

Thus, we can estimate the energy change and texture variation of the HR image from the LR image according to Eqs. (3) and (4), respectively. In the next subsection, we will introduce how to predict the visual quality of HR images based on these two components.

2.4. Overall Quality Prediction

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

$$Q = F_e^\beta F_t^\gamma \quad (5)$$

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; β and γ are parameters used to adjust the relative importance of these two components. Here, we regard the factors of energy change and texture variation as the same important and set β and γ as one.

3. EXPERIMENTAL RESULTS

We use the database with subjective scores in [5] to do the comparison experiment. Although the ground truth HR images are available in [5], 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 [20], where the images are with diverse content obtained in a professional photographic style. The SR images are generated from LR images by six existing single frame super-resolution algorithms. In total, there are 540 SR images 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 to 10 for each SR image.

Table 2. Performance evaluation of the proposed method.

	PSNR	SSIM	MSSSIM	NQM	VIF	MAD	NSS-SR	Proposed
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

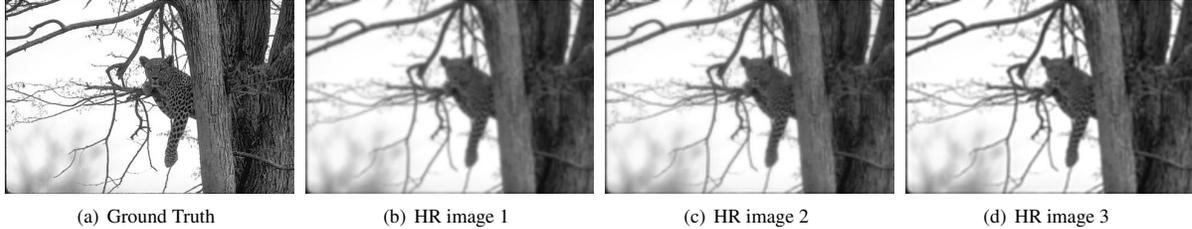


Fig. 2. The ground truth image (a) and HR images. (b) MOS: 1.8462, RRIQA-SR:0.4397, MSSSIM: 0.8414, NQM: 17.6, PSNR: 19.3077; (c) MOS: 2.0769, RRIQA-SR: 0.4831, MSSSIM: 0.8737, NQM: 22.3425, PSNR: 19.9234; (d) MOS: 2.1538, RRIQA-SR:0.4853, MSSSIM: 0.8631, NQM: 18.7231, PSNR: 19.7341.

The subjective perceptual quality of HR images is represented by the mean opinion score (MOS). Kendall rank-order correlation coefficient (KRCC) and Spearman rank-order correlation coefficient (SRCC) are used to evaluate the performance of different quality metrics. Generally, a better IQA metric can obtain higher KRCC and SRCC values.

We first analyze the performance of the proposed method on these two components in Eq. (5): the energy change F_e , and the texture variation F_t . Experimental results are listed in Table 1. From this table, we can see that the evaluation results from F_t can obtain higher correlation with subjective data than those from F_e , which 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.

To further demonstrate the performance of the proposed method, we use the NSS-SR metric [14, 16] designed specifically for image super-resolution/interpolation to conduct the comparison experiment. The following full reference quality metrics are also used in performance evaluation due to the available ground truth information: PSNR, SSIM [12], multi-scale SSIM (MSSSIM) [21], noise quality measure (NQM) [23], visual information fidelity (VIF) [24], and the most apparent distortion (MAD) [26]. Experimental results are shown in Table 2.

From Table 2, we can see that MSSSIM can obtain better performance than SSIM and PSNR, similar with visual quality for general images [21]. The reason is that the MSSSIM uses more high-frequency information for quality prediction. NQM and MAD can obtain better performance in

quality prediction than VIF and MSSSIM. 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 can be well measured by NQM and MAD. From Table 2, NSS-SR obtains the lowest performance among the compared quality metric. Although NSS-SR is designed for image super-resolution, the NSS models used in that metric are built specifically for image interpolation [14, 16]. The used HR images are created by using various image super-resolution algorithms rather than image interpolation [5]. Thus, the NSS-SR can not work well in this database.

In Fig. 2, we provide some HR image samples with different scores calculated from some different metrics. From this figure, we can see that all used quality metrics can predict the consistent quality of HR images 1 and 2 with subjective data. However, for HR image 3 with better quality than HR image 2, all used metrics (including MSSSIM, NQM and PSNR) cannot predict its quality well. In contrast, the proposed RRIQA-SR can predict the visual quality of all these images consistently with the subjective data.

4. 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. The visual quality of HR images is predicted by energy change and texture variation in HR images, which are computed based on the similarity between image patches of LR and HR image patches. The experimental results show that the proposed RRIQA-SR method can obtain better performance than other quality metrics, even some full-reference quality metrics. In the future, we will investigate how to use the proposed RRIQA-SR to optimize image super-resolution algorithms.

5. REFERENCES

- [1] M. Shen, C. Wang, P. Xue, and W. Lin, Performance of Reconstruction-based Super-resolution with Regularization, *Journal of Visual Communications and Image Representation*, Vol.21(7), pp. 640-650, 2010.
- [2] C. Wang, P. Xue, and W. Lin, Improved Super-resolution Reconstruction from Video, *IEEE Trans. CSVT*, vol.16(11), pp.1411-1422, 2006.
- [3] R. G. Keys, Cubic convolution interpolation for digital image processing, *IEEE Trans. Asoustics, Speech, and Signal Processing*, 29, 1153-1160, 1981.
- [4] R. Fattal, Image upsampling via imposed edge statistics, *ACM TOG*, 26(3): 95, 2007.
- [5] C.-Y. Yang, C. Ma, and M.-H. Yang, Single-image super-resolution: a benchmark, *ECCV*, 2014.
- [6] Y. Zhang, J. Liu, W. Yang, and Z. Guo, Image super-resolution based on structure-modulated sparse representation, *IEEE TIP*, 24(9), 2797-2810, 2015.
- [7] Y.-Q. Zhang, Y. Ding, J. Liu, and Z. Guo, Guided image filtering using signal subspace projection, *IET Image Process.*, 7(3), pp. 270-279, 2013.
- [8] K. Ni, and T. Nguyen, Image superresolution using support vector regression, *IEEE TIP*, 16(6), 1596-1610, 2007.
- [9] J. Yang, J. Wright, T. S. Huang, and Y. Ma, Image super-resolution via sparse representation, *IEEE TIP*, 19(11), pp. 2861-2873, 2010.
- [10] C. Dong, C. C. Loy, K. He, and X. Tang, Learning a deep convolutional network for image super-resolution, *ECCV*, 2014.
- [11] W. Lin, and C.-C. Jay Kuo. Perceptual Visual Quality Metrics: A Survey. *Journal of Visual Communication and Image Representation*, 22(4): 297- 312, 2011.
- [12] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE TIP*, Vol. 13(4), 600-612, 2004.
- [13] J. Wu, W. Lin, G. Shi, and A. Liu. Perceptual Quality Metric With Internal Generative Mechanism. *IEEE TIP*, 22(1): 43-54, 2013.
- [14] H. Yeganeh, M. Rostami and Z. Wang, Objective quality assessment for image super-resolution: a natural scene statistics approach, *IEEE ICIP*, 2012.
- [15] A. R. Reibman, R. M. Bell, and S. Gray, Quality assessment for super-resolution image enhancement, *IEEE ICIP*, 2006.
- [16] H. Yeganeh, M. Rostami and Z. Wang, Objective quality assessment of interpolated natural images, *IEEE TIP*, 24(11), 4651-4663, 2015.
- [17] N. Ahmed, T. Natarajan, K. R. Rao, Discrete Cosine Transform, *IEEE Transactions on Computers*, C-23(1): 90-93, 1974.
- [18] R. C. Gonzalez, and R. E. Woods, *Digital Image Processing (3rd Edition)*, Prentice Hall, 2008.
- [19] C. Y. Yang, M. H. Yang, Fast direct super-resolution by simple functions, *IEEE ICCV*, 2013.
- [20] D. Martin, C. Fowlkes, D. Tal, and J. Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, *IEEE ICCV*, 2001.
- [21] Z. Wang, E. Simoncelli, and 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.
- [22] H. R. Sheikh, A. C. Bovik, and G. de Veciana, An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE TIP*, 14(12), 2117-2128, 2005.
- [23] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, Image quality assessment based on a degradation model, *IEEE TIP*, 9(4), 636650, 2000.
- [24] H. R. Sheikh, and A. C. Bovik, Image information and visual quality, *IEEE TIP*, 15(2): 430-444, 2006.
- [25] L. Li, W. Lin, X. Wang, G. Yang, K. Bahrami, and A. C. Kot, No-Reference Image Blur Assessment Based on Discrete Orthogonal Moments. *IEEE Trans. Cybernetics*, 46(1), 39-50, 2016.
- [26] E. C. Larson and D. M. Chandler, Most apparent distortion: fullreference image quality assessment and the role of strategy, *Journal of Electronic Imaging*, vol. 19, p. 011006, 2010.
- [27] S. Z. Li, *Markov Random Field Modeling in Image Analysis*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2001.
- [28] C. Liu, J. Yuen and A. Torralba. SIFT flow: dense correspondence across different scenes and its applications. *IEEE T-PAMI*, 33(5), 2011.
- [29] D. G. Lowe. Object recognition from local scale-invariant features. *IEEE ICCV*, 1999.