Pamukkale Univ Muh Bilim Derg, 27(5), 610-617, 2021

Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi

Pamukkale University Journal of Engineering Sciences

Image quality assessment based on manifold distortion

Manifold bozulması ile imge kalitesi değerlendirme

Mehmet TÜRKAN[1*]

[1]Department of Electrical and Electronics Engineering, Engineering Faculty, Izmir University of Economics, Izmir, Turkey.
mehmet.turkan@ieu.edu.tr

Received/Geliş Tarihi: 04.06.2020 Revision/Düzeltme Tarihi: 10.11.2020 doi: 10.5505/pajes.2020.69158
Accepted/Kabul Tarihi: 17.11.2020 Research Article/Araştırma Makalesi

Abstract

An image quality metric is proposed by introducing a new framework for full reference image quality assessment from the perspective of image patch manifolds. Assuming that most natural scenes are sampled from low dimensional manifolds or submanifolds, perceived image degradations in structural variations can be quantitatively evaluated on the surfaces of highly nonlinear image manifolds. Manifold distortion image quality index first characterizes intrinsic geometric properties of the locally linear manifold structures of spatially local patch spaces, and then measures the deviation from the original smooth manifold structure to calculate the distortion index. Experimental results demonstrate a strong promise with a comparison to both subjective evaluation and state-of-the-art objective quality assessment methods.

Keywords: Image quality assessment, Image quality index, Manifold learning, Neighbor embedding.

Öz

Görüntü parçacık manifoldları perspektifinden, yeni bir tam referans görüntü kalitesi değerlendirmesi çerçevesi oluşturularak bir görüntü kalitesi metriği önerilmektedir. Çoğu doğal sahnenin düşük boyutlu manifoldlardan veya alt-manifoldlardan örneklendiği varsayılarak, yapısal varyasyonlarda algılanan görüntü bozulmaları yüksek derecede doğrusal olmayan görüntü manifoldlarının yüzeylerinde nicel olarak değerlendirilebilir. Manifold bozulması görüntü kalite endeksi önce uzamsal olarak yerel parçacık uzaylarının yerel doğrusal manifold yapılarının içsel geometrik özelliklerini karakterize etmekte ve daha sonra bozulma endeksini hesaplamak için orijinal pürüzsüz manifold yapısından sapmayı ölçmektedir. Deneysel sonuçlar hem öznel değerlendirme hem de gelişmiş objektif kalite değerlendirme yöntemleriyle kıyaslandığında güçlü bir taahhüt göstermektedir.

Anahtar kelimeler: Görüntü kalite değerlendirmesi, Görüntü kalite endeksi, Manifold öğrenmesi, Komşuluk gömülmesi.

1 Introduction

One of the most important research topics in image and video processing is the quality assessment (QA) of a visual content. During and after the acquisition, any captured image or video may have been introduced different kinds of distortion until it has been projected or presented to a human observer. There are basically two main sources of distortion. The first source is related to the hardware limitations where characteristics of sensing elements, transmission channels and display devices play an important role in the final display quality. The second one is related to the software tools especially along with the content editing and compression, as well as transmission and storage algorithms applied to the visual content. Since any kinds of captured and processed visual content are devoted for a final human consumption, it is obviously very crucial to imitate the human visual system (HVS) and obtain objective quality metrics in an excellent agreement with subjective opinion from human observers. Therefore, the principal aim is to design generic quantitative QA models which are highly aligned with the HVS, to accurately estimate the perceptual visual quality of a visual content.

Depending on the availability of a reference (i.e., pristine original) content, image and video QA methods can be categorized into three groups. In the first group, there is only the distorted image available and the evaluation of the perceived quality of this image must be done without any reference, i.e., no-reference QA. The second group which is referred to reduced-reference QA, on the other hand, contains partial information about the reference content in addition to the distorted image. This type QA methods makes use of the available partial knowledge of the reference to assess the quality of the distorted content. The third group consists in a full-reference image and a distorted version of it where the distortion level is measured based on a comparison with the available distortion-free reference. In this study, the focus will be on full-reference QA algorithms which basically serve a great basis for designing distortion resilient image/video processing applications, e.g., for acquisition, communication, compression, editing, displaying and printing. Moreover, an objective full-reference quality metric will be a key component of embedded systems evaluating image fidelity or perceived similarity.

Conventional full-reference QA algorithms, including mean squared error (MSE) [1] and its extensions, i.e., signal-to-noise ratio (SNR) and peak SNR (PSNR), are mainly based on low level models of the HVS. Although MSE is an inexpensive signal fidelity measure with its simple mathematical convention, it can only estimate point-by-point errors in pixel domain, or in an appropriate transform domain, without considering structural changes between the distortion-free and distorted images. Weighted SNR (WSNR) [2],[3] can yet be thought as an extension to MSE in which an appropriate weighting has been incorporated in the distortion measure. The weighting strategy of WSNR employs a contrast sensitivity function that is a linear spatially invariant approximation of the HVS. Nonlinear noise quality measure (NQM) [4] respects to the nonlinear spatially varying characteristics of the HVS. NQM simply behaves as a nonlinear weighted SNR. Universal image quality index (UQI) [5] assumes that the HVS is highly adapted to extract structural

*Corresponding author/Yazışılan Yazar

information from a scene. It models any image distortion with a combination of three measures: loss of correlation, luminance distortion and contrast distortion. UQI has been generalized to structural similarity (SSIM) [6], and SSIM has later been extended to multiscale SSIM (MSSIM) [7] and information content weighted SSIM (IWSSIM) [8]. Most apparent distortion (MAD) [9] combines visual detection and image appearance-based distortion models to measure the overall perceived distortion. Gradient similarity metric (GSM) [10] measures structural and contrast changes using the gradient similarity in images. It also incorporates a luminance similarity for a complete quality assessment. In addition, statistical approaches assume that there are parallels between the models of natural scenes (visual stimulus of the natural environment) and the evolution of the HVS [11],[12]. Information fidelity criterion (IFC) [11] models the scale invariant statistics of images in the wavelet domain. Visual information fidelity (VIF) [12] can be regarded as a content dependent extension of IFC. Visual SNR (VSNR) [13] also operates in a multiscale wavelet domain. Furthermore, there are feature based similarity metrics based on transform features, e.g., RFSIM [14], phase congruency and gradient magnitude features, e.g., FSIM [15], local binary patterns [16] and image descriptors, e.g., SURF-SIM [17].

This paper develops a novel framework for full-reference QA from the perspective of image manifolds. The proposed solution builds upon a (multi)patch-based scheme which naturally leads to a constrained optimization problem for characterizing the intrinsic properties of image manifold structures to measure a local distortion index. Experimental results illustrate a strong promise when compared to both subjective evaluation and objective QA methods. The rest of this paper is organized as follows. Section 2 introduces the motivation and details the main ideas and steps of the proposed QA algorithm. Section 3 presents the experimental setup and draws a discussion over the obtained results. Section 4 finally gives a brief conclusion while describing possible future directions.

2 Manifold distortion quality index

A very well studied approach in image and video processing is to make full use of textural and structural self-similarities within an image and cross-similarities across different images. The fact is that sufficiently small image patches are very likely to repeat themselves within the same scale and across different scales of an image, as well as across different images. Based on this observation, most of the image processing problems such as inpainting [18]-[21], denoising [22]-[24], super-resolution [25]-[28], image prediction and compression [29]-[35], and more have found their solutions in the literature. All these methods have mainly been inspired from texture synthesis techniques [36]-[38] which are highly influenced by Markov Random Fields [39],[40].

A second observation suggests that most of the natural images are sampled from low-dimensional (sub)manifolds. This assumption leads to the fact that densely sampled and sufficiently small image texture patches can be successfully reconstructed as a weighted linear combination of their nearest-neighboring (NN) patches. This is usually referred to as neighbor-embedding [41]-[44] and is basically aligned with the main idea of manifold learning algorithms for dimensionality reduction [45]-[47].

In this work, the texture synthesis principle and the manifold sampling assumption have been combined resulting in a new framework for full-reference QA. Given a pair of images of same size, i.e., the pristine original and its distorted version, a patch based scheme builds upon a strictly localized NN search followed by a characterization of the intrinsic properties of the manifold structures of spatially local patch spaces, and then by a calculation of the projected local deviation from the smooth distortion-free manifold structure, in order to calculate the distortion index per pixel.

2.1 Problem definition and notation

Given two images of same size $N \times M$ pixels represented by $\mathcal{X}$ as the distortion free original and $\mathcal{Y}$ as the distorted version of $\mathcal{X}$, the main objective is to design a quantitative QA model in order to accurately estimate the perceptual visual quality of $\mathcal{Y}$ using the information contained in $\mathcal{X}$. As stated above, the manifold sampling assumption will be followed and extended for this purpose. The solution naturally leads to a local multi-patch scheme which is a powerful and generic enough tool to deal with different kinds of distortion corruption in $\mathcal{Y}$.

A lexical ordering of image pixels and $n \times n$ square image patches as stacked column vectors of size $n^2 \times 1$ are assumed. While representing the given images as matrices by $\mathbf{X}$ and $\mathbf{Y}$, each distortion-free patch $\mathbf{x}_i$ extracted from $\mathbf{X}$ has a spatially collocated distorted patch $\mathbf{y}_i$ obtained from $\mathbf{Y}$. Here column vectors $\mathbf{x}_i$ and $\mathbf{y}_i$ in fact denote $n \times n$ square patches centered around the pixel indexed by i extracted from $\mathbf{X}$ and $\mathbf{Y}$, respectively, where $i = 1 \dots NM$ and n is odd. Let us denote these patch pairs in a set as $\mathcal{S} = \{\mathbf{x}_i, \mathbf{y}_i\}_{\forall i}$.

2.2 Manifold distortion as a quality index

The main objective here is to characterize point-based image distortions by means of patch (sub)manifolds through spatially close and chromatically similar image texture patches. In order to achieve this aim, distortion-free local neighborhood information of $\mathbf{x}_i \ \forall i$ has first been extracted from a strictly localized search region Ω_i centered around $\mathbf{x}_i$, by minimizing a weighted distance metric $\|\mathbf{x}_i - \mathbf{x}_j\|_{2,\sigma}^2$ where σ represents the standard deviation of the $n \times n$ Gaussian kernel used for weighting and $\{\mathbf{x}_j\} = \{\mathbf{x}_i \mid \mathbf{x}_i \in \Omega_i \wedge i \neq j\}$. K-closest such neighbors are kept in a set $\{\mathbf{x}_{ik}\}_{k=1}^K$ as the K-NN of $\mathbf{x}_i$. Then similar to [46], intrinsic local geometric properties of each individual neighborhood can be linearly characterized as in Equation (1) by solving

$$\underset{\{\alpha_{ik}\}}{\operatorname{argmin}} \left\| \mathbf{x}_i - \sum\nolimits_k \alpha_{ik} \mathbf{x}_{ik} \right\|_{2,\sigma}^2 \quad s.t. \quad \sum\nolimits_k \alpha_{ik} = 1, \forall i. \tag{1}$$

The above constrained least squares optimization can be solved for $\boldsymbol{\alpha}_i = [\alpha_{i1}\ \alpha_{i2} \dots \alpha_{iK}]^T$ as given in Equation (2) by

$$\boldsymbol{\alpha}_i = \frac{\mathbf{G}_{i,\sigma}^{-1}\mathbf{1}}{\mathbf{1}^T \mathbf{G}_{i,\sigma}^{-1}\mathbf{1}} \tag{2}$$

where $\mathbf{G}_{i,\sigma} = \overline{\mathbf{X}}_{i,\sigma}^T \overline{\mathbf{X}}_{i,\sigma}$ is the Gram (inner product) matrix and the columns of the matrix $\overline{\mathbf{X}}_{i,\sigma}$ correspond to the set $\{\bar{\mathbf{x}}_{ik}^{\sigma}\}$ representing the $\mathbf{x}_i$-centered and then Gaussian weighted NN set $\{\mathbf{x}_{ik}\}$, and $\mathbf{1}$ denotes a column vector of ones of size $K \times 1$. The optimum reconstruction weights $\{\alpha_{ik}\}_{k=1}^K$ here describe the local properties of the distortion-free manifold structure.

Now let us think of a special case when there is no distortion on the pristine original content, i.e., $\mathcal{X} = \mathcal{Y}$, the local neighborhood of and the manifold structure around $\mathbf{x}_i$ would be identical for

$\mathbf{y}_i$ $\forall i$. This fact leads to a fundamental observation that, these local geometric similarities of image spaces can be benefited from, and a quality index can be successfully calculated based on manifold distortions, by relating the intrinsic properties of distortion-free and distorted local patch neighborhoods via neighbor-embedding. To do so, the parameters of distortion-free local neighborhood information is directly transferred to the distorted image manifold structure as if there were no distortion. These parameters simply correspond to the index set $\{ik\}$ of the selected K-NN of $\mathbf{x}_i$ which will help extract collocated distorted image patches $\{\mathbf{y}_{ik}\}_{k=1}^{K}$ from the correct, indeed ground-truth spatial coordinates in $\mathbf{Y}$. After transferring this local neighborhood information to the distorted domain, intrinsic local geometric properties of distorted neighborhood of $\mathbf{y}_i$ can be characterized as in Equation (3) by solving

$$\underset{\{\omega_{ik}\}}{\operatorname{argmin}} \left\| \mathbf{y}_i - \sum\nolimits_k \omega_{ik}\mathbf{y}_{ik} \right\|_{2,\sigma}^2 \quad s.t. \quad \sum\nolimits_k \omega_{ik} = 1, \forall i, \tag{3}$$

for $\boldsymbol{\omega}_i = [\omega_{i1}\ \omega_{i2} \dots \omega_{iK}]^T$ as given in Equation (4) by

$$\boldsymbol{\omega}_i = \frac{\mathbf{G}_{i,\sigma}^{-1}\mathbf{1}}{\mathbf{1}^T\mathbf{G}_{i,\sigma}^{-1}\mathbf{1}}. \tag{4}$$

$\mathbf{G}_{i,\sigma} = \bar{\mathbf{Y}}_{i,\sigma}^T\bar{\mathbf{Y}}_{i,\sigma}$ is the inner product matrix where the columns of the matrix $\bar{\mathbf{Y}}_{i,\sigma}$ correspond to the set $\{\bar{\mathbf{y}}_{ik}^{\sigma}\}$ representing the $\mathbf{y}_i$-centered and Gaussian weighted NN set $\{\mathbf{y}_{ik}\}$, and $\mathbf{1}$ denotes a column vector of respective size. The optimum reconstruction weights $\{\omega_{ik}\}_{k=1}^{K}$ describe the desired local properties of the distorted manifold structure.

The information extracted above is for both local (point-based via Gaussian weighting) and nonlocal (patch-based via reconstruction weights) intrinsic characteristics of distortion-free and distorted local neighborhoods, and thus, is a valuable tool for measuring textural and structural deviations from the original distortion-free structure. While transferring distorted local neighborhood parameters set $\{\omega_{ik}\}$ directly to the original manifold structure from which quantitative deviations can easily be measured, two different representations of $\mathbf{x}_i$ $\forall i$ are reconstructed using the distortion-free local neighborhood $\{\mathbf{x}_{ik}\}$. The first one is the distortion-free representation of $\mathbf{x}_i$, i.e., $\bar{\mathbf{x}}_i = \sum_k \alpha_{ik}\mathbf{x}_{ik}$; and the second holds the properties of the distorted manifold structure, i.e., $\bar{\bar{\mathbf{x}}}_i = \sum_k \omega_{ik}\mathbf{x}_{ik}$. The next step then consists in measuring the distortion between $\bar{\mathbf{x}}_i$ and $\bar{\bar{\mathbf{x}}}_i$ as $\mathbf{d}_i = \bar{\mathbf{x}}_i - \bar{\bar{\mathbf{x}}}_i = \sum_k(\alpha_{ik} - \omega_{ik})\mathbf{x}_{ik}$. Figure 1 summarizes all these parameter transfer procedures between distortion-free and distorted images and corresponding local manifold structures.

During local neighborhood selection and characterization by the least-squares optimization, much more attention has been given to the centering pixel indexed by i. Thus, the **M**anifold **D**istortion **Q**uality **I**ndex can now be defined as $\mathbf{MDQI}[i] = P(\mathbf{d}_i)$. The operator P first reshapes $\mathbf{d}_i$ into an $n \times n$ distortion patch, then extracts the value of the centering pixel location and clamps it to $[-2^m + 1, 2^m - 1]$ for m-bit images, if necessary. Finally, a Manifold Distortion MSE (MDMSE) and a Manifold Distortion PSNR (MDPSNR) are formulated in Equation (5) by

$$\mathbf{MDMSE} = \frac{1}{|\mathcal{S}|}\sum_i (\mathbf{MDQI}[i])^2$$
$$\mathbf{MDPSNR} = 20\log_{10}\frac{2^m - 1}{\sqrt{\mathbf{MDMSE}}} \tag{5}$$

where $|\mathcal{S}|$ denotes the number of patch pairs in the set $\mathcal{S}$.

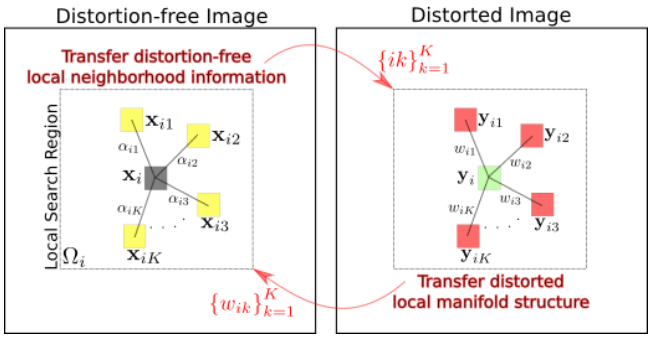

Figure 1. Parameter transfer between distortion-free and distorted images and corresponding manifolds.

3 Experimental setup and results

The Tampere Image Database 2013 (TID2013) [48] and The LIVE Image Quality Assessment Database (LIVE, Release 2) [49],[50] are utilized and tested for the proposed QA algorithm. TID2013 contains 25 true color (24-bits/pixel RGB) distortion-free and 3000 distorted images. In this database, 24 types of distortions with 5 different levels are introduced, i.e., there are 120 distorted images per reference image. The LIVE Image Database contains 29 true color distortion-free images which are distorted by 5 types of distortions with different perceptual quality levels. LIVE consists of a total 982 test images (including 203 reference images). The distortion types and details of these datasets are given in Table 1.

Table 1. The distortion details of TID2013 and LIVE datasets.

Database	Type	Description
TID	AGN	additive Gaussian noise
	ANC	additive noise in color components is more intensive than additive noise in luminance
	SCN	spatially correlated noise
	MN	masked noise
	HFN	high frequency noise
	IN	impulse noise
	QN	quantization noise
	GB	Gaussian blur
	ID	image denoising
	JP1	JPEG compression
	JP2K1	JPEG2K compression
	JP2	JPEG transmission errors
	JP2K2	JPEG2K transmission errors
	NEPN	non eccentricity pattern noise
	LBD	local block-wise distortions of different intensity
	MS	mean (intensity) shift
	CC	contrast change
	CCS	change of color saturation
	MGN	multiplicative Gaussian noise
	CN	comfort noise
	LCNI	lossy compression of noisy images
	CQD	image color quantization with dither
	CA	chromatic aberrations
	SSR	sparse sampling and reconstruction
LIVE	JPG2K	JPEG2K compression (227 images)
	JPG	JPEG compression (233 images)
	GBlur	Gaussian blur (174 images)
	WN	white noise in the RGB components (174 images)
	FF	JPEG2K transmission errors in the bit-stream using a fast-fading Rayleigh channel (174 images)

TID2013 and LIVE also provide the mean opinion score (MOS) for each distorted image. In TID2013, a total of 971 observers from Finland (116), France (72), Italy (80), Ukraine (602) and USA (101) have evaluated relative visual quality in 1048680 image pairs, i.e., 524340 visual quality comparisons of distorted images. The obtained MOS values lie between 0 (minimal) and 9 (maximal). The higher MOS value is the better visual quality of the image. In LIVE, about 20-29 human subjects have rated visual quality of each image in seven different experiments, using the same equipment and viewing conditions. Observers have evaluated each distortion type in order to provide a visual quality on a continuous linear scale with five equal regions labeled as "Bad", "Poor", "Fair", "Good" and "Excellent". These raw scores have later been translated into a difference MOS value for each distorted image.

It is worth noting here that a single scale of analysis might not be effective since both image resolution and viewing distance have great effect on the perceived quality. In practice, it has been suggested to use an empirical formula to determine the scale for images viewed from a typical distance [6]. In this study, all images are automatically decimated by a factor $F = \max(1, \text{round}(\min(M, N)/256))$ with a simple averaging filter of size $F \times F$ pixels. Another keynote is that the elements of the set □ can be used to calculate the distortion index for all pixels, or a subset of pixels in the image for the sake of computational complexity. In the latter case, a subset $\tilde{\mathcal{S}}$ can be chosen from □ with a predetermined regular pixel offsets in both dimensions of the image. This paper assumes $\tilde{\mathcal{S}} = \mathcal{S}$.

In order to evaluate and compare the performance of MDQI with the competing methods, four widely utilized metrics are employed, namely Spearman rank-order correlation coefficient (SROCC) [51], Kendall rank-order correlation coefficient (KROCC) [52], Pearson linear correlation coefficient (PLCC) and root MSE (RMSE). The calculation details of these metrics are illustrated in Table 2.

Table 2. The details of SROCC, KROCC, PLCC and RMSE for a given paired data $\{(a_1, b_1), \ldots, (a_N, b_N)\}$ consisting of N pairs.

Metric	Formula	Description
SROCC	$1 - \frac{6\sum_n d_n^2}{N(N^2-1)}$	d_n: difference between two ranks of each observation, $n = 1 \ldots N$
KROCC	$\frac{2(N_c - N_d)}{N(N-1)}$	N_c= number of rank-concordant pairs N_d= number of rank-discordant pairs
PLCC	$\frac{\sum_n (a_n - \mu_a)(b_n - \mu_b)}{\sqrt{\sum_n (a_n - \mu_a)^2}\sqrt{\sum_n (b_n - \mu_b)^2}}$	μ_a= sample mean of a μ_b= sample mean of b $n = 1 \ldots N$
RMSE	$\sqrt{\frac{\sum_n (a_n - b_n)^2}{N}}$	$n = 1 \ldots N$

While SROCC and KROCC measure the ranking monotonicity between the subjective MOS and the objective score variables, both PLCC and RMSE need a prior regression analysis to get a nonlinear mapping between these variables. For this purpose, a five-parameter logistic function with a linear term that is constrained to be monotonic is utilized as in Equation (6) by

$$f(x) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\beta_2 (x - \beta_3)}} \right) + \beta_4 x + \beta_5 \tag{6}$$

where the parameter set $\{\beta_p\}_{p=1}^{5}$ needs to be fitted to minimize the MSE between the mapped objective scores and the MOS [49]. A better objective QA metric is expected to have higher SROCC, KROCC and PLCC values, but a lower RMSE value.

For computational purposes, local search region Ω_i is fixed to a square window of size 27×27 pixels centered around the pixel indexed by i $\forall i$, and all image patches are of size 9×9 pixels, i.e., $n = 9$. This block size has been experimentally shown to be large enough in order to capture local structures and fine details, to be small enough in order to apply neighbor-embedding in terms of texture synthesis. The standard deviation σ of the 9×9 Gaussian kernel used for weighting is set to 3.50, and then this kernel is uniformly normalized to have a centering coefficient equal to "1". The same kernel is applied both for local neighborhood (K-NN) selection and for local geometry characterization by least-squares optimization. K is set to 8 providing a sparsity notion for the representation. It is important note here that all these parameters are fixed in the reported experimental results for a fair comparison with the competing methods.

For distortion-free and distorted neighborhood optimizations, mean subtracted image patches are employed as textural features rather than patch absolute intensities. Although first and/or second order gradients could be extracted as in [41] and [53] for this purpose, it is sufficient here to consider representative features as relative textural and structural variations with respect to the patch mean. In this way, the distortion $\mathbf{d}_i$ can easily detect local intensity and structural changes.

Figure 2 depicts a visual example of image quality index maps for the *Einstein* image. In this example, all distorted images have roughly the same MSE values with respect to the original image, but they have obviously different visual quality. MDMSE leads to a better indication of image quality assessment together with a visual index map. The details of the calculated QA statistics with MDQI in comparison to the widely used MSE and SSIM scores are summarized in Table 3.

All experiments performed for all quality metrics on TID 2013 and LIVE are reported in Table 4, Table 5, and Table 6. Table 4 and Table 5 illustrate detailed performance comparison on each individual distortion type (as listed in Table 1) in terms of SROCC on TID 2013 and LIVE, respectively. In these tables, the statistics are calculated on each distortion type separately and the top-three performing QA models are given in bold (per-row). It can be clearly observed from these statistics, MDQI demonstrates a strong global promise when compared to both state-of-the-art objective methods and subjective evaluation scores. In addition to that, it is better than the most distortion types, and in some cases a comparable alternative to the other QA models. On the other hand, MDQI tends to fail when there are local or global uniform intensity changes and/or contrast changes between images. This can be well-explained with the constrained optimization of each local neighborhood, which is in fact translation invariant because of the sum-to-one constraint on weights. This observation however is aligned with the HVS which is insensitive to uniform intensity changes up to a level.

Table 6 further gives an overall performance comparison of several QA models in terms of SROCC, KROCC, PLCC and RMSE on TID 2013 and LIVE datasets. In this table, the statistics are calculated on the whole datasets employed and the top-three performing QA models are highlighted in bold (per-row). MDQI outperforms most of the models on TID 2013 and in some cases, it is comparable to the other QA models on LIVE.

The nonlinear regression parameters in Equation (6) are obtained using iterative least squares estimation with initial

values set to $\{\beta_p\} = \{\text{MOS}_\sigma, 1, \text{QA}_\mu, 1, 0.1\}$ where MOS_σ and QA_μ represent the standard deviation of the MOS and the mean of the QA model selected, respectively. This initialization of parameters gives a meaningful conditioning while reducing the probability of having a local minimum in the final estimation.

Figure 3 demonstrates some examples of scatter plots of the MOS versus MDQI together with other selected QA models, i.e., IWSSIM, GSM, MAD, FSIM, MSSIM, SSIM, PSNR. The black curves shown in this figure are obtained by the logistic function. It can also be seen from these scatter plots that MDQI has a consistent correlation with the MOS.

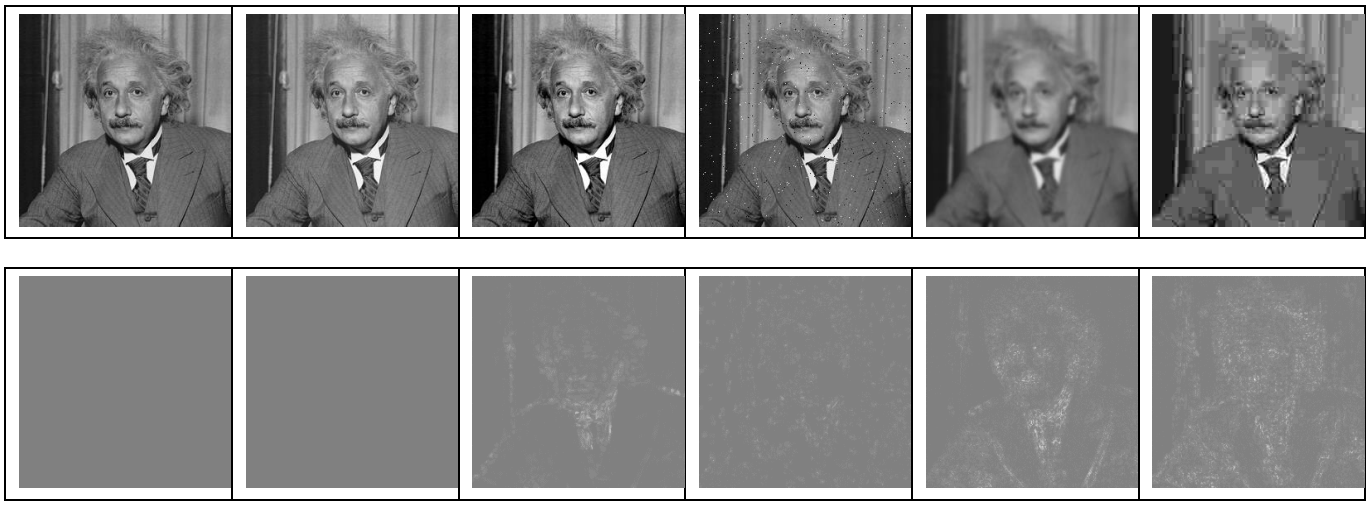

Figure 2. The *Einstein* image. (Top Left-to-right) Original (MSE = 0, SSIM = 1, MDMSE = 0); uniform mean shift (MSE = 144, SSIM = 0.988, MDMSE = 0.001); contrast change (MSE=144, SSIM = 0.913, MDMSE = 30.50); impulse noise (MSE = 144, SSIM = 0.840, MDMSE=20.93); blur (MSE=144, SSIM=0.694, MDMSE = 96.41); JPEG compression (MSE = 142, SSIM = 0.662, MDMSE = 121.48). (Bottom Left-to-right) MDQI maps of the original, uniform mean shift, contrast change, impulse noise, blur and JPEG compression.

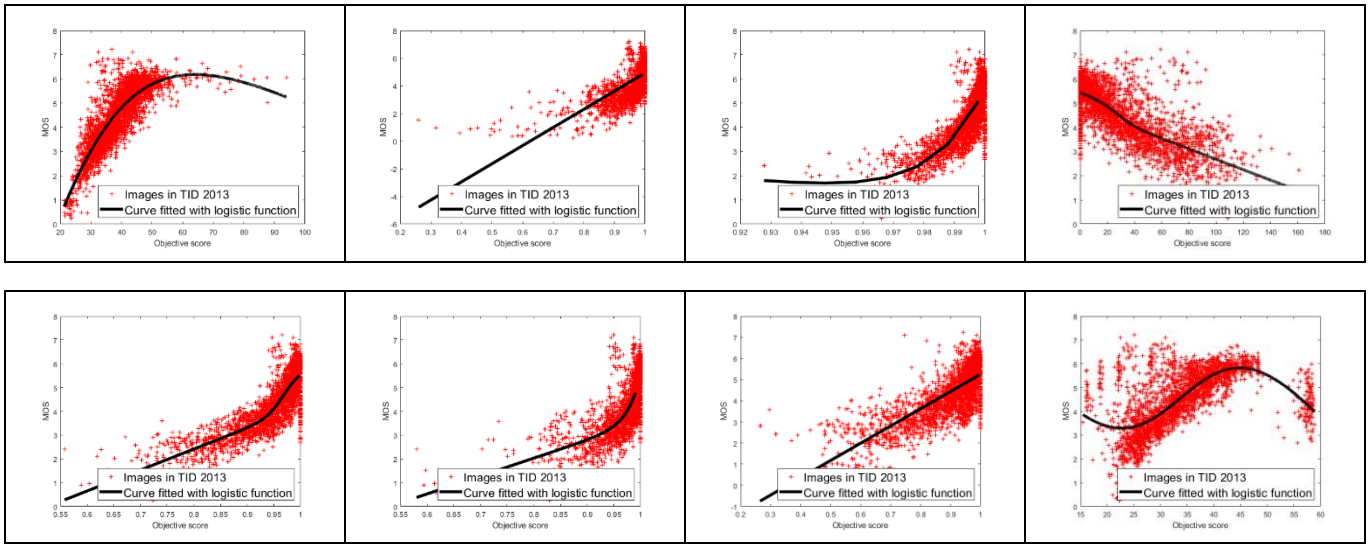

Figure 3. Example scatter plots of the MOS vs. scores of QA models on TID 2013. The black curves are obtained by the logistic regression function. (Top Left-to-right) MDQI, IWSSIM, GSM, MAD; (Bottom Left-to-right) FSIM, MSSIM, SSIM, PSNR.

Table 3. The *Einstein* image statistics.

Type	MSE	SSIM	MDMSE
Original	0	1	0
Mean Shift	144	0.988	0.001
Contrast Change	144	0.913	30.50
Impulse Noise	144	0.840	20.93
Blur	144	0.694	96.41
JPEG Compression	142	0.662	121.48

Table 4. Performance comparison of several QA models for each individual distortion type of TID2013 in terms of SROCC.

TID	PSNR	SSIM	MSSIM	VSNR	VIF	UQI	IFC	NQM	WSNR	SNR	FSIM	MAD	GSM	IWSSIM	IWPSNR	MDQI
AGN	**0.936**	0.867	0.861	0.824	0.914	0.583	0.707	0.814	0.855	0.864	0.897	0.859	0.906	0.858	**0.927**	**0.916**
ANC	**0.891**	0.773	0.769	0.664	**0.848**	0.496	0.589	0.718	0.811	0.838	0.818	0.738	0.818	0.766	**0.864**	0.831
SCN	**0.920**	0.851	0.854	0.807	0.908	0.606	0.724	0.787	0.882	0.850	0.875	0.871	**0.916**	0.846	**0.956**	**0.916**
MN	**0.799**	0.777	0.745	0.662	**0.837**	0.567	0.691	0.690	0.595	0.718	0.794	0.573	0.729	0.754	0.681	**0.814**
HFN	**0.950**	0.863	0.869	0.814	0.907	0.657	0.773	0.863	0.892	**0.915**	0.897	0.850	0.887	0.874	**0.918**	**0.915**
IN	**0.899**	0.750	0.750	0.790	0.870	0.505	0.622	0.791	**0.909**	0.898	0.808	0.687	0.796	0.745	**0.914**	0.888
QN	**0.904**	0.866	0.871	0.854	0.861	0.656	0.662	0.826	**0.890**	0.879	0.871	0.866	0.884	0.859	**0.898**	0.854
GB	0.950	**0.967**	0.962	0.945	**0.966**	0.912	0.905	0.901	0.934	0.936	0.955	0.863	**0.969**	0.964	0.917	0.957
ID	**0.951**	0.925	0.931	0.921	0.927	0.767	0.826	0.918	0.920	**0.938**	0.930	0.906	**0.943**	0.925	0.914	0.935
JP1	**0.960**	0.920	0.920	0.860	**0.937**	0.786	0.869	0.887	0.919	0.927	0.933	0.894	0.928	0.917	0.935	**0.948**
JP2K1	**0.969**	0.947	0.948	0.921	0.958	0.866	0.910	0.926	0.927	0.924	0.958	0.913	**0.960**	0.951	0.951	**0.970**
JP2	0.798	**0.849**	0.824	0.755	**0.856**	0.845	0.773	0.736	0.665	0.771	0.847	0.756	**0.851**	0.820	0.768	0.827
JP2K2	**0.950**	0.883	0.874	0.820	0.890	0.749	0.809	0.808	0.826	**0.897**	0.891	0.860	**0.918**	0.865	0.857	**0.918**
NEPN	0.733	0.782	0.804	0.736	0.812	0.769	0.542	0.747	0.788	0.703	0.792	**0.828**	**0.813**	0.811	0.804	**0.830**
LBD	0.087	**0.572**	0.144	0.110	0.497	0.513	0.420	0.001	0.138	0.071	**0.551**	0.059	**0.642**	0.185	0.081	0.426
MS	0.767	0.775	**0.791**	0.574	0.625	0.744	0.580	0.608	0.773	0.774	0.752	0.598	**0.787**	**0.789**	0.617	0.606
CC	0.431	0.377	0.461	0.334	**0.828**	0.361	0.358	0.462	0.418	0.422	**0.468**	0.192	**0.486**	0.451	0.419	0.246
CCS	0.009	**0.414**	0.362	0.159	0.230	0.301	**0.379**	0.124	0.077	0.006	0.378	0.038	0.359	0.340	0.088	**0.679**
MGN	**0.898**	0.780	0.780	0.801	0.875	0.505	0.666	0.768	0.800	**0.878**	0.847	0.793	0.835	0.794	**0.889**	**0.889**
CN	**0.936**	0.857	0.878	0.858	0.923	0.738	0.835	0.870	0.916	0.890	0.912	0.874	0.912	0.892	**0.931**	**0.944**
LCNI	0.948	0.906	0.910	0.914	0.949	0.780	0.850	0.905	0.949	0.907	0.947	0.932	**0.956**	0.912	**0.967**	**0.958**
CQD	**0.928**	0.854	0.882	0.882	0.877	0.665	0.669	0.862	**0.896**	**0.897**	0.876	0.864	**0.897**	0.877	0.893	0.885
CA	**0.889**	0.877	0.868	0.847	0.859	0.783	0.804	0.810	0.827	**0.883**	0.871	0.766	**0.882**	0.859	0.820	0.848
SSR	**0.962**	0.946	0.950	0.940	0.957	0.867	0.918	0.946	0.948	0.919	0.956	0.947	**0.967**	0.953	0.959	**0.964**
avg	0.811	0.807	0.792	0.741	**0.838**	0.668	0.703	0.740	0.773	0.779	0.826	0.730	**0.835**	0.792	0.790	**0.832**
std	0.257	**0.149**	0.193	0.225	0.163	**0.159**	0.160	0.231	0.235	0.249	**0.149**	0.258	**0.148**	0.190	0.246	0.174

Table 5. Performance comparison of several QA models for each individual distortion type of LIVE in terms of SROCC.

LIVE	PSNR	SSIM	MSSIM	VSNR	VIF	UQI	IFC	NQM	WSNR	SNR	FSIM	MAD	GSM	IWSSIM	IWPSNR	MDQI
JPG2K	0.970	0.984	**0.985**	0.967	**0.988**	0.934	0.955	0.973	0.966	0.957	**0.988**	0.957	**0.987**	0.983	0.983	0.977
JPG	0.970	**0.984**	**0.985**	0.961	**0.987**	0.950	0.965	0.979	0.977	0.965	**0.987**	0.952	**0.985**	**0.984**	**0.984**	0.971
GBlur	0.931	0.972	0.980	0.973	**0.981**	0.968	0.970	0.908	0.919	0.913	**0.983**	0.956	0.972	**0.984**	0.965	0.952
WN	**0.989**	0.982	0.987	0.988	**0.992**	0.934	0.966	**0.991**	0.981	0.984	0.980	0.978	0.987	0.988	0.986	0.985
FF	0.946	**0.974**	0.931	0.905	0.959	0.961	0.961	0.896	0.896	0.948	**0.971**	0.945	**0.965**	0.941	0.874	0.947
avg	0.961	**0.979**	0.974	0.959	**0.981**	0.949	0.963	0.949	0.948	0.953	**0.982**	0.958	**0.979**	0.976	0.958	0.966
std	0.020	**0.005**	0.021	0.028	0.012	0.014	**0.005**	0.039	0.034	0.023	**0.006**	0.011	**0.009**	0.018	0.043	0.015

Table 6. Performance comparison of several QA models. TID2013 and LIVE in terms of SROCC, KROCC, PLCC and RMSE.

TID	PSNR	SSIM	MSSIM	VSNR	VIF	UQI	IFC	NQM	WSNR	SNR	FSIM	MAD	GSM	IWSSIM	IWPSNR	MDQI
SROCC	0.634	0.742	0.756	0.663	0.769	0.633	0.570	0.646	0.533	0.617	**0.802**	0.721	**0.795**	0.742	0.630	**0.832**
KROCC	0.493	0.559	0.579	0.496	0.592	0.458	0.422	0.476	0.404	0.467	**0.629**	0.549	**0.626**	0.563	0.465	**0.647**
PLCC	0.705	0.760	0.805	0.689	0.824	0.691	0.673	0.676	0.611	0.681	**0.859**	0.760	**0.846**	0.670	0.602	**0.852**
RMSE	0.880	0.806	0.735	0.899	0.702	0.897	0.917	0.913	0.981	0.908	**0.635**	0.806	**0.660**	0.921	0.991	**0.650**
LIVE	**PSNR**	**SSIM**	**MSSIM**	**VSNR**	**VIF**	**UQI**	**IFC**	**NQM**	**WSNR**	**SNR**	**FSIM**	**MAD**	**GSM**	**IWSSIM**	**IWPSNR**	**MDQI**
SROCC	0.958	0.973	0.970	0.957	**0.979**	0.940	0.952	0.954	0.953	0.952	**0.981**	0.957	**0.977**	0.973	0.952	0.965
KROCC	0.832	0.864	0.862	0.829	**0.883**	0.802	0.810	0.830	0.827	0.821	**0.888**	0.844	**0.876**	0.873	0.825	0.844
PLCC	0.949	0.828	0.637	0.956	**0.965**	0.880	0.879	0.949	0.946	0.941	0.848	**0.961**	0.772	0.651	0.945	**0.960**
RMSE	9.86	17.53	24.12	**9.17**	**8.25**	14.86	14.94	9.86	10.14	10.59	16.60	**8.71**	19.89	23.75	10.21	**8.71**

4 Conclusion

In this study, a novel technique for full-reference QA through manifold learning is developed with results highly correlated to subjective assessments. To the best of available knowledge, the proposed algorithm is a new framework which takes advantage of texture synthesis and manifold sampling through neighbor embeddings of image patches. The proposed model, namely MDQI, i.e., MDMSE and MDPSNR, is capable of quantitively evaluating perceived image degradations in structural and textural variations, and moreover it can produce an index map because of its pixel-based structure. Possible future directions include the investigation of other efficient ways to measure distortion values through intrinsic properties of each individual neighborhood, a gradient-features based characterization of neighborhoods, an adaptation of the constrained optimization to handle local/global uniform intensity and contrast changes, and a multi-scale extension to MDQI (MMDQI) not only for increasing its accuracy but also for assessing quality of images with different sizes.

5 Author contribution statements

In the scope of this study, Mehmet TÜRKAN, in the formation of the idea, the design and the literature review, performing analyzes and examining the results, the spelling and checking the article in terms of content were contributed.

6 Ethics committee approval and conflict of interest statement

There is no need to obtain permission from the ethics committee for the article prepared.

There is no conflict of interest with any person / institution in the article prepared.

7 References

[1] Wang Z, Bovik AC. "Mean squared error: Love it or leave it? A new look at signal fidelity measures". *IEEE Signal Processing Magazine*, 26(1), 98-117, 2009.

[2] Mannos J, Sakrison D. "The effects of a visual fidelity criterion of the encoding of images". *IEEE Transactions on Information Theory*, 20(4), 525-536, 1974.

[3] Mitsa T, Varkur KL. "Evaluation of contrast sensitivity functions for the formulation of quality measures incorporated in halftoning algorithms". *IEEE 1993 International Conference on Acoustics, Speech, and Signal Processing*, Minneapolis, MN, USA, 27-30 April 1993.

[4] Damera-Venkata N, Kite TD, Geisler WS, Evans BL, Bovik AC. "Image quality assessment based on a degradation model". *IEEE Transactions on Image Processing*, 9(4), 636-650, 2000.

[5] Wang Z, Bovik AC. "A universal image quality index". *IEEE Signal Processing Letters*, 9(3), 81-84, 2002.

[6] Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. "Image quality assessment: From error visibility to structural similarity". *IEEE Transactions on Image Processing*, 13(4), 600-612, 2004.

[7] Wang Z, Simoncelli EP, Bovik AC. "Multiscale structural similarity for image quality assessment". *2003 Asilomar Conference on Signals, Systems & Computers*, Pacific Grove, CA, USA, 9-12 November 2003.

[8] Wang Z, Li Q. "Information content weighting for perceptual image quality assessment". *IEEE Transactions on Image Processing*, 20(5), 1185-1198, 2011.

[9] Larson EC, Chandler DM. "Most apparent distortion: Full-reference image quality assessment and the role of strategy". *Journal of Electronic Imaging*, 2010. https://doi.org/10.1117/1.3267105.

[10] Liu A, Lin W, Narwaria M. "Image quality assessment based on gradient similarity". *IEEE Transactions on Image Processing*, 21(4), 1500-1512, 2012.

[11] Sheikh HR, Bovik AC, de Veciana G. "An information fidelity criterion for image quality assessment using natural scene statistics". *IEEE Transactions on Image Processing*, 14(12), 2117-2128, 2005.

[12] Sheikh HR, Bovik AC. "Image information and visual quality". *IEEE Transactions on Image Processing*, 15(2), 430-444, 2006.

[13] Chandler DM, Hemami SS. "VSNR: A wavelet-based visual signal-to-noise ratio for natural images". *IEEE Transactions on Image Processing*, 16(9), 2284-2298, 2007.

[14] Zhang L, Zhang L, Mou X. "RFSIM: A feature based image quality assessment metric using Riesz transforms". *IEEE 2010 International Conference on Image Processing*, Hong Kong, China, 26-29 September 2010.

[15] Zhang L, Zhang L, Mou X, Zhang D. "FSIM: A feature similarity index for image quality assessment". *IEEE Transactions on Image Processing*, 20(8), 2378-2386, 2011.

[16] Wu J, Lin W, Shi G. "Image quality assessment with degradation on spatial structure". *IEEE Signal Processing Letters*, 21(4), 437-440, 2014.

[17] Wang F, Sun X, Guo Z, Huang Y, Fu K. "An object-distortion based image quality similarity". *IEEE Signal Processing Letters*, 22(10), 1534-1537, 2015.

[18] Criminisi A, Perez P, Toyama K. "Region filling and object removal by exemplar-based image inpainting". *IEEE Transactions on Image Processing*, 13(9), 1200-1212, 2004.

[19] Zhang Y, Xiao J, Shah M. "Region completion in a single image". *2004 Eurographics*, Grenoble, France, 30 August-3 September 2004.

[20] Sun J, Yuan L, Jia J, Shum HY. "Image completion with structure propagation". *ACM Transactions on Graphics*, 24(3), 861-868, 2005.

[21] C. Barnes, Shechtman E, Finkelstein A, Goldman DB. "PatchMatch: A randomized correspondence algorithm for structural image editing". *ACM Transactions on Graphics*, 2009. https://doi.org/10.1145/1531326.1531330.

[22] Buades A, Coll B, Morel J. "A non-local algorithm for image denoising". *IEEE 2005 Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, CA, USA, 20-25 June 2005.

[23] Mahmoudi M, Sapiro G. "Fast image and video denoising via nonlocal means of similar neighborhoods". *IEEE Signal Processing Letters*, 12(12), 839-842, 2005.

[24] Dabov K, Foi A, Katkovnik V, Egiazarian K. "Image denoising with block-matching and 3D filtering". *SPIE 2006 Electronic Imaging*, San Jose, CA, USA, 17 February 2006.

[25] Freeman WT, Jones TR, Pasztor EC. "Example-based super-resolution". *IEEE Computer Graphics and Applications*, 22(2), 56-65, 2002.

[26] Glasner D, Bagon S, Irani M. "Super-resolution from a single image". *IEEE 2009 International Conference on Computer Vision*, Kyoto, Japan, 29 September-2 October 2009.

[27] Freedman G, Fattal R. "Image and video upscaling from local self-examples". *ACM Transactions on Graphics*, 30(2), 2011. https://doi.org/10.1145/1944846.1944852.

[28] Michaeli T, Irani M. "Nonparametric blind super-resolution". *IEEE 2013 International Conference on Computer Vision*, Sydney, NSW, Australia, 1-8 December 2013.

[29] Sugimoto K, Kobayashi M, Suzuki Y, Kato S, Boon CS. "Inter frame coding with template matching spatio-temporal prediction". *IEEE 2004 International Conference on Image Processing*, Singapore, Republic of Singapore, 24-27 October 2004.

[30] Yang J, Yin B, Sun Y, Zhang N. "A block-matching based intra frame prediction for H.264/AVC". *IEEE 2006 International Conference on Multimedia and Expo*, Toronto, Ontario, Canada, 9-12 July 2006.

[31] Tan TK, Boon CS, Suzuki Y. "Intra prediction by template matching". *IEEE 2006 International Conference on Image Processing*, Atlanta, GA, USA, 8-11 October 2006.

[32] Tan TK, Boon CS, Suzuki Y. "Intra prediction by averaged template matching predictors". *IEEE 2007 Consumer Communications and Networking Conference*, Las Vegas, NV, USA, 11-13 January 2007.

[33] Turkan M, Guillemot C. "Sparse approximation with adaptive dictionary for image prediction". *IEEE 2009 International Conference on Image Processing*, Cairo, Egypt, 7-10 November 2009.

[34] Turkan M, Guillemot C. "Image prediction: Template matching vs. sparse approximation". *IEEE 2010 International Conference on Image Processing*, Hong Kong, China, 26-29 September 2010.

[35] Turkan M, Guillemot C. "Image prediction based on neighbor-embedding methods". *IEEE Transactions on Image Processing*, 21(4), 1885-1898, 2012.

[36] Efros AA, Leung TK. "Texture synthesis by non-parametric sampling". *IEEE 1999 International Conference on Computer Vision*, Kerkyra, Greece, 20-27 September 1999.

[37] Wei LY, Levoy M. "Fast texture synthesis using tree-structured vector quantization". *2000 Annual Conference on Computer Graphics and Interactive Techniques*, New Orleans, LA, USA, 23-28 July 2000.

[38] Ashikhmin M. "Synthesizing natural textures". *2001 Symposium on Interactive 3D Graphics*, Chapel Hill, NC, USA, 26-29 March 2001.
[39] Besag J. "Spatial interaction and the statistical analysis of lattice systems". *Journal of the Royal Statistical Society Series B*, 36(2), 192-236, 1974.
[40] Cross GR, Jain AK. "Markov random field texture models". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5(1), 25-39, 1983.
[41] Chang H, Yeung DY, Xiong Y. "Super-resolution through neighbor embedding". *IEEE 2004 Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, DC, USA, 27 June-2 July 2004.
[42] Turkan M, Thoreau D, Guillotel P. "Self-content super-resolution for ultra-HD up-sampling". *2012 European Conference on Visual Media Production*, London, UK, 5-6 December 2012.
[43] Turkan M, Thoreau D, Guillotel P. "Optimized neighbor embeddings for single-image super-resolution". *IEEE 2013 International Conference on Image Processing*, Melbourne, VIC, Australia, 15-18 September 2013.
[44] Turkan M, Thoreau D, Guillotel P. "Iterated neighbor-embeddings for image super-resolution". *IEEE 2014 International Conference on Image Processing*, Paris, France, 27-30 October 2014.
[45] Tenenbaum JB, de Silva V, Langford JC. "A global geometric framework for nonlinear dimensionality reduction". *Science*, 290(5500), 2319-2323, 2000.
[46] Roweis ST, Saul LK. "Nonlinear dimensionality reduction by locally linear embedding". *Science*, 290(5500) 2323-2326, 2000.
[47] Donoho DL, Grimes C. "Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data". *2003 Proceedings of the National Academy of Sciences*, 100(10), 5591-5596, 2003.
[48] Ponomarenko N, Jin L, Ieremeiev O, Lukin V, Egiazarian K, Astola J, Vozel B, Chehdi K, Carli M, Battisti F, Kuo CCJ. "Image database TID2013: Peculiarities, results and perspectives". *Signal Processing: Image Communication*, 30, 57-77, 2015.
[49] Sheikh HR, Sabir MF, Bovik AC. "A statistical evaluation of recent full reference image quality assessment algorithms". *IEEE Transactions on Image Processing*, 15(11), 3440-3451, 2006.
[50] Sheikh HR, Wang Z, Cormack L, Bovik AC. "LIVE Image Quality Assessment Database Release 2". http://live.ece.utexas.edu/research/quality (03.05.2020).
[51] Spearman C. "The proof and measurement of association between two things". *The American Journal of Psychology*, 15(1), 72-101, 1904.
[52] Kendall MG. "A new measure of rank correlation". *Biometrika*, 30(1-2), 81-93, 1938.
[53] Yang J, Wright J, Huang TS, Ma Y. "Image super-resolution via sparse representation". *IEEE Transactions on Image Processing*, 19(11), 2861-2873, 2010.