Reduced-reference quality assessment of multiply-distorted images based on structural and uncertainty information degradation

Mahmoudpour, Saeed; Schelkens, Peter

Published in:
Journal of Visual Communication and Image Representation

DOI:
10.1016/j.jvcir.2018.10.027

Publication date:
2018

J. Vis. Commun. Image R. 57 (2018) 125–137

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Reduced-reference quality assessment of multiply-distorted images based on structural and uncertainty information degradation

Saeed Mahmoudpour [a,b,*], Peter Schelkens [a,b]

[a] Vrije Universiteit Brussel (VUB), Department of Electronics and Informatics, Pleinlaan2, B-1050 Brussels, Belgium
[b] imec, Kapeldreef 75, B-3001 Leuven, Belgium

ARTICLE INFO

Article history:
Received 5 April 2018
Revised 1 October 2018
Accepted 27 October 2018
Available online 29 October 2018

Keywords:
Image quality
Multiply-distortion types
Shearlet transform
Entropy analysis
Support vector regression
Privileged information

ABSTRACT

The majority of existing objective Image Quality Assessment (IQA) methods are designed for evaluation of images corrupted by single distortion types. However, images may be degraded with multiple distortions during processing stages. In this paper, we propose a reduced-reference IQA algorithm to predict the quality of multiply-distorted images. An image is first decomposed into predicted and disorderly portions based on the internal generative mechanism theory. The structural information is captured from the predicted image by using a shearlet representation and Rényi directional entropy is deployed to measure the disorderly information changes. Finally, we introduce the application of a framework namely Learning Using Privileged Information (LUPI) to build a quality model and obtain quality scores. During training, the LUPI framework utilizes a set of additional privileged data to learn an improved quality model. Experimental results on multiply-distorted image datasets (MLIVE and MDID2015) confirm the effectiveness of the proposed IQA model.

1. Introduction

With the tremendous growth of multimedia technology and advances in image acquisition devices, digital images have become more popular and prevalent in our daily life. Despite of this progress, the quality of digital images may be degraded since they are subjected to various distortions in end-to-end application chains. Therefore, it is crucial to evaluate and maintain the visual quality.

Image Quality Assessment (IQA) is an active research area in recent years. Subjective IQA has been known as the most reliable method of quality assessment but it is tedious and cumbersome. Hence, objective IQA metrics are developed that use computational models for predicting the visual quality [1–3].

Depending on the accessibility to reference/pristine image, the objective IQA methods can be categorized into three types: Full-Reference (FR), Reduced-Reference (RR) and No-Reference (NR). FR IQA metrics [4,5] often have better performance than the two other IQA types while they require access to the entire reference image for quality prediction. RR methods [6,7] use only limited information of the reference image and NR IQA algorithms [8,9] predict image quality without using any reference data. Gu et al. [10] proposed a NR metric for quality assessment of enhanced images. The method extracts quality-relevant features by analysing different image attributes (such as contrast, brightness, sharpness, and etc.) and the quality score is pooled through a robust model learned from a massive number of training samples. The application of NR IQA for evaluation of screen content pictures is investigated in [11]. Since natural images are often richer in colour and include more structures as well as higher luminance range than screen content images, the authors devised specific features to better represent the characteristics of screen content images. A blind IQA method is proposed in [12] to address the quality prediction of Depth Image-Based Rendering (DIBR) synthesized images. To capture the geometric distortion introduced during DIBR, the algorithm exploits error between the DIBR image and its autoregressive (AR) predicted image.

FR IQA is impractical in many applications where the undistorted reference image is not available. Due to a variety of image contents and distortion types, it is very challenging to design effective NR IQA metrics that can perform well without using any reference data. RR IQA algorithms attain a good trade-off between FR and NR approaches as they employ limited information from the reference image while delivering a better performance.

In general, RR IQA models are based on extracting a set of quality-sensitive features from the reference and distorted images.

* Corresponding author at: Vrije Universiteit Brussel (VUB), Department of Electronics and Informatics, Pleinlaan2, B-1050 Brussels, Belgium.

E-mail addresses: smahmoud@etrovub.be (S. Mahmoudpour), Peter.Schelkens@vub.be (P. Schelkens).

https://doi.org/10.1016/j.jvcir.2018.10.027
1047-3203/

Wang et al. [13] developed a natural image statistic model in the wavelet domain for RR quality evaluation. The distance between the marginal probability distributions of the wavelet coefficients is measured between the reference and test images to obtain a quality score. According to the Orientation Selectivity (OS) mechanism of the primary visual cortex, Wu et al [14] proposed an OS-based Visual Pattern (OSVP) to acquire visual features for RR IQA. Soundararajan et al. [7] proposed an RR approach that uses the difference of the weighted entropies between the reference and distorted images in the wavelet domain. Rehman et al. [15] developed an RR-IQA algorithm based on the principle of structural similarity utilizing the statistical properties of the image in the Divisive Normalization Transform (DNT) domain. Zhang et al. [6] quantify local sharpness of images to develop an RR IQA index. The authors classify distorted images into eight families based on generated sharpness maps and then deploy a regression framework on sharpness features to predict class-specific quality scores.

Most of the existing IQA algorithms are designed for quality prediction of images suffering from single distortion types. However, multiple distortions may be introduced to images in various processing stages. Digital images may pass through acquisition, compression and transmission steps before reaching end users. Thus, images may be subjected simultaneously to multiple distortion types hence cluttering final IQA. Chandler et al. [16] mentioned some of the possible joint effects and interferences between different distortion types and pointed out that multiply-distortion IQA is one of the most important challenges to be solved. Several multiply-distorted image datasets have been released. The first broadly-used established datasets are MDID2013 [17] and MLIVE [18] that consist of images corrupted by three types of mixed distortions. A new dataset namely MDID2015 (multiply-distorted image dataset) [19] has been established recently, which contains images subjected simultaneously to upto 4 types of distortions. The multiply-distorted image datasets challenge many state-of-art IQA methods and it is of great importance to design IQA models suitable for multiply-distorted images. When multiple distortions influence an image, the joint degradation effect may not be well interpreted by studying the effect of the contributed single distortions, because individual distortions can affect each other and produce different degradation characteristics. One type of distortion may have a masking effect on the other type or the distortions may combine in a way that complicates the perception and intensify the visual quality degradation.

Gu et al. [20] proposed an NR IQA metric for multiply-distorted images which consists of a de-noising step followed by sharpness and blockiness degree estimation. The final score is obtained by pooling the blur and blockiness scores. These authors also developed an improved method [17] by considering the effect of each emerging single distortion type (noise, JPEG and blur) together with their joint effects. An NR IQA method is proposed in [21] based on feature extraction from a gradient weighted histogram of using Local Binary Patterns (LBP) in which the proposed structural features can suitably describe the complex degradation pattern introduced by multiple distortions.

In this paper, we propose an RR method based on Shearlets and Entropy Analysis (RRSEA) for quality assessment of multiply-distorted images. The approach is inspired by the facts that: (i) the Human Visual System (HVS) attempts to minimize the free-energy of the observed scene by predicting the main visual content and ignoring the residual uncertainty [19,20]. Based on this, an input image is decomposed into predicted and disorderly (residual uncertainty) parts [21]. (ii) Multiple types of distortions have a joint effect on both predicted and disorderly parts (i.e. on primary visual information and uncertainty data) of an image. Here, shearlet coefficients and directional entropy are utilized to measure the effect of distortion on predicted and disorderly parts, respectively. We have reported an initial version of this work and its preliminary results at the 2017 QoMEX conference [22]. This paper provides more elaborated experimental validation in which the proposed method proved competitive results with respect to several state-of-the-art metrics. Moreover, to train our model, we introduce the application of an improved machine learning model namely Learning Using Privileged Information (LUPI) in the field of image quality assessment. To the best of our knowledge, no such work has been reported in the literature. In the following paragraphs, we further elaborate on the concept of the proposed RRSEA IQA method.

Recent findings in the field of brain and cognitive sciences reveal the Internal Generative Mechanism (IGM) [23,24] of the HVS. Based on the IGM theory, HVS tries to predict the main visual structures and avoid the residual uncertainty. Wu et al. [25] proposed a model based on IGM that decomposes an image into predicted and disorderly portions. Because the predicted part contains the primary visual information and the disorderly part holds the uncertainty data, the effect of a specific type of distortion is not similar on both portions and it depends on the inherent characteristics of that distortion. In [26], an NR IQA framework is developed that extracts three types of features based on free energy principle, HVS-inspired structural information, and Natural Scene Statistics (NSS)-model to predict quality score. A blind sharpness metric is proposed in [27] upon the AR and energy principle models. The method quantifies sharpness by percentile pooling of local energy and contrast differences on AR estimated coefficients. Liu et al. [28] developed a RR metric based on free energy principle that utilize a sparse representation to provide a prediction of reference and distorted images. Finally, the discrepancies between two predictions are quantified to pool a quality score.

The interaction of multiple distortion types brings a more complex problem and complicate the design of an effective metric [18,29]. There have been advances in quality evaluation of multiply-distorted images in recent years; however, there still remains significant room for improvement. The existing approaches are either limited in the number of measured distortion types [20,17] or they only consider the degradation of the main structural information with less emphasis on the joint effects of different distortions [21]. In this work, we suggest to use the IGM-based decomposition for better analysis of the joint effect of multiply-distortion types. Particularly, we simultaneously consider the distinct influence of a multiply-distortion type on main visual/structural information located in the predicted part and uncertainty information in disorderly part.

The basic idea of RRSEA is to quantify the statistical property deviations occurred in natural images when a distortion is applied. First, an input image is decomposed into predicted and disorderly parts according to the IGM theory. To obtain effective IQA features, the structural information in the predicted part is extracted by a shearlet transform [30]. The shearlets can provide accurate representation of structural data by analysing input predicted signal at multiple directions. Therefore, the structural variations introduced by distortions can be well captured. The Rényi directional entropy [31] is measured for modeling the uncertainty information in the disorderly part. Image entropy presents the amount of information in images. Since the distortions lead to significant information changes in all directions, we deployed an entropy method with directional selectivity to effectively capture the disorderly data changes in natural images. The extracted features from both predicted and disorderly parts are concatenated as a single feature vector to capture the joint effect of multiply-distortion on two portions. In the quality prediction stage, we utilized a new Support Vector Regression [26] (SVR)-based paradigm called LUPI to build a model and predict the quality score. Vapnik et al. [32] introduced LUPI to improve the predictive performance of learning algorithms using additional information (called privileged measures). Such privileged data are only accessible during training and not in testing. The SVM-based realization of the LUPI paradigm is called SVM + for classification and the regression realization is denoted as ε–SVR+.

The LUPI paradigm has recently gained significant attention in machine learning community [33,34]. Yang et al. [35] applied the paradigm in face detection where some additional facial features such as head poses and gender were utilized only during training. Sarafianos et al. [28] applied the LUPI for human height estimation in surveillance systems. The anthropometric measurements that are relatively difficult to be estimated in surveillance cameras (e.g., circumferences of human parts) are referred as privileged information and they are used only during training to learn a better classifier.

During training, LUPI uses some additional FR information (privileged measures) together with the RR information (obtained from our IGM-based method). The goal is to exploit such privileged measures during training phase to learn an improved quality model. In testing stage, the privileged measures are not available and the image quality is predicted using the RR information and the trained model.

The experimental results show that the constructed model based on ε–SVR+ can improve the IQA performance and the proposed method outperforms several IQA approaches. Summarized, this paper brings the following contributions: (1) a new RR IQA is proposed for quality evaluation of multiply-distorted images (2) the LUPI paradigm is employed to learn an improved quality prediction model.

The remainder of the paper is organized as follows: The basic introduction of the IGM theory and the mathematical description of the image decomposition method are explained in Section 2. Section 3 presents the feature extraction for quality assessment from structural and uncertainty information. Section 4 elaborates on learning the IQA models based on the LUPI paradigm. The experimental results are presented and discussed in Section 5. Finally, Section 6 concludes this paper.

2. IGM-based image decomposition

Recent investigations in brain sciences reveal that the brain provides a probabilistic representation of information and accordingly, mathematical concepts such as Bayesian theory [36] and the free energy principle [23] are introduced. These concepts indicate that the Human Visual System (HVS) possesses an Internal Generative Mechanism (IGM) [37].

According to IGM, the brain behaves as an active predictor when analyzing a scene. For visual processing, the brain first tries to generate predictions of the visual stimuli by detecting correlations of nearby contents while ignoring uncertainty data. Then, the predictions are merged with other subjective attributes, most notably inherent a priori knowledge, to optimize the active prediction. The HVS depicts a specific functionality when exploring a scene. The visual system is well-adapted to extract homogeneous and regular contents (such as structures) to understand the scene. On the other hand, determining the composition of irregular parts is not straightforward for the HVS and the disorder information brings difficulty for visual information prediction and understanding. In other words, the prediction of a visual scene is highly linked with the similarities among nearby contents [38].

Based on the aforementioned characteristics, IGM-based prediction is simplified by considering the correlation of neighbouring pixels. A Bayesian prediction-based autoregressive (AR) model is proposed in [25] to predict a pixel value by computing its correlation with nearby pixels. The AR model imitates the visual content prediction of IGM and decomposes an image into predicted and disorderly portions. Using the AR model, the value $p_c$of a central pixel x_c can be predicted as:

$$p_c = \sum_{p_i \in X} \widehat{R}_i . p_i + \varepsilon \quad (1)$$

where p_c is the predicted value of x_c and p_i is the value of neighborhood pixels x_i $(x_i \in X)$. The normalized correlation coefficient between x_c and x_i is represented by $\widehat{R}_i = \frac{R_i}{\sum_k R_k}$ and ε is a term characterizing the white noise. The correlation coefficient is more likely to be high for pixels located in parts with regular structures and homogeneous content while the pixels of disorderly regions have lower correlation.

The predicted pixels obtained by (1) constitute the predicted image (I_P). Therefore, the image I is decomposed into its predicted I_P and disorderly $I_D(= I - I_P)$ portions. Fig. 1a shows a reference image with its Blur-JPEG and Blur-Noise multiply-distorted versions. The images are selected from the MLIVE dataset [18] and their corresponding decomposed parts are also presented. Fig. 1b shows the enlarged versions of the regions marked in Fig. 1a. The main visual content, e.g. edges and structures, is represented in the predicted image and the disorderly portion conveys the residual uncertainty information. Distortions on the predicted part mainly affect the image structure and subsequently the image understanding while disorderly part distortions mostly modify the image disorder causing an uncomfortable perception with limited impact on visual understanding.

The degradation effects of different distortion types are not similar for the predicted/disorderly portions and depend on the characteristics of a distortion type. For instance, Gaussian Blur (GB) can eliminate structures as well as the uncertainty information. The blur deteriorates visual understanding by modifying the primary visual information (e.g. edges) and texture (or uncertainty information). Oppositely, noise causes uncomfortable perception, impacting mostly the disorderly portion and it does not have a significant effect on image structure. A Blur-Noise distorted image and its decomposed parts are shown in Fig. 1. It is observed that the blur mainly damages the predicted portion whereas the noise affects more the disorderly portion. For Blur-JPEG multiply-distorted images, blur mostly degrades the structures in the predicted part and JPEG distortion affects both parts with less effect on the predicted part. Since each distortion type has a distinct impact on the predicted and disorderly portions, we proposed to utilize the image decomposition inspired by IGM for quality prediction of multiply-distorted images.

It is challenging to design an effective IQA model for multiply-distorted images since the interaction between several distortions should be taken into account. We advocate that the degradation effect of multiple distortions can be better interpreted when evaluating both the predicted and disorderly images. The structural degradations emerge on the predicted part while the disorderly part represents the information changes that disturb the perception with a slight effect on visual understanding. Using the IGM-based prediction model, a new RR IQA approach is developed to quantify the degradation effect of multiply-distortion types on two decomposed parts and predict the perceived quality.

3. Proposed method

The framework of the proposed method is summarized in Fig. 2. The method consists of three main stages including: image decomposition, feature extraction, and quality prediction (feature pooling). The input train/test images are first decomposed into predicted and disorderly parts. In the next step, some quality-characterizing features are extracted from each portion. The fea-

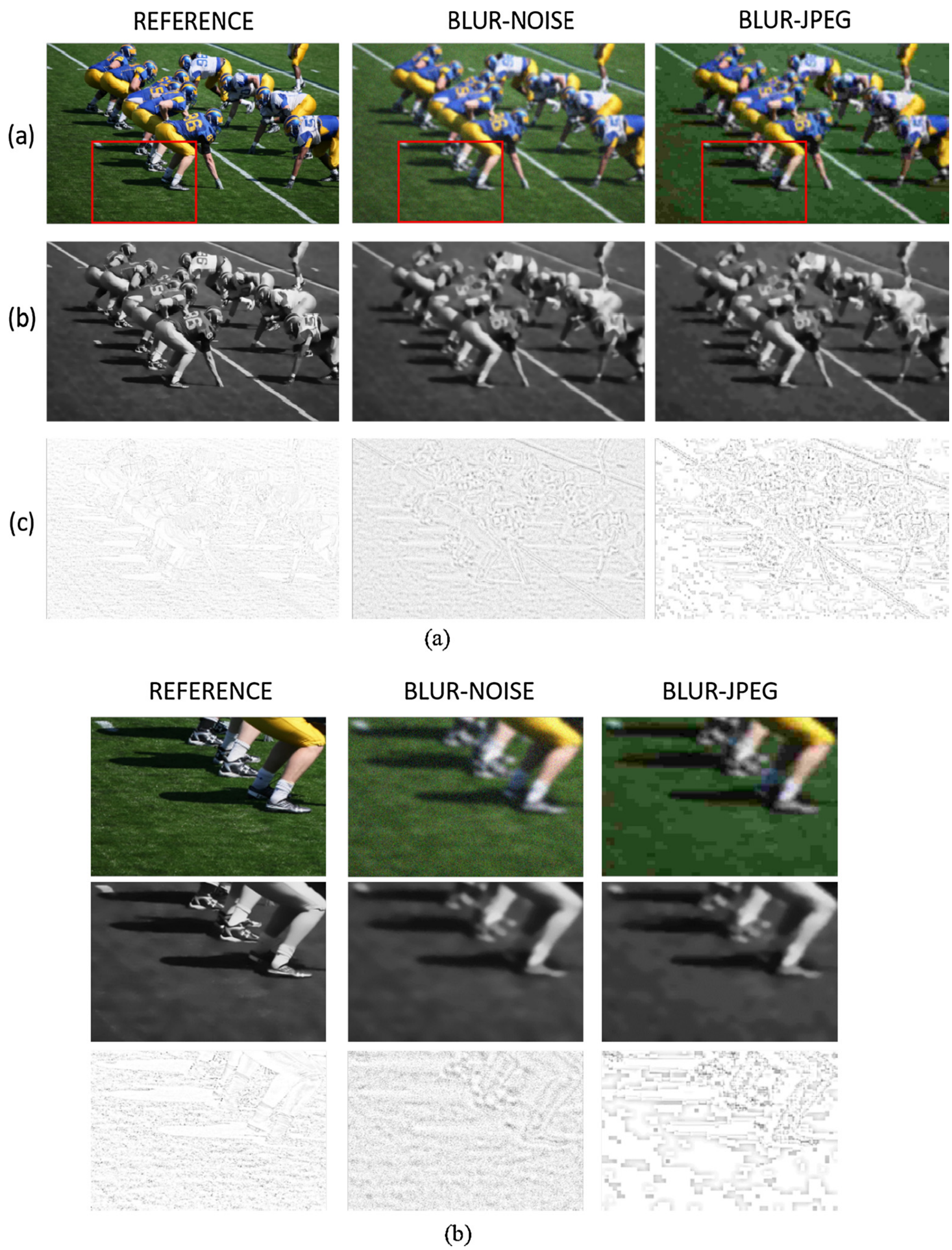

Fig. 1. Image decomposition based on IGM (*scrimmage.bmp* from MLIVE dataset). (a) First row: Input images (Reference image and the Blur-Noise and Blur-JPEG distorted versions), Second row: predicted part of each image (I_P), Third row: disorderly part of each image (I_D) [scaled to 0–255]. (b) Zoom-in versions of the regions marked by red box in (a) and their corresponding predicted and disorderly parts.

tures of the predicted part are obtained using a shearlet representation and the directional entropies of the disorderly part are computed yielding a number of disorderly features. The difference between features of the reference and distorted images are then computed to obtain RR quality measures. In the training stage, the obtained measures are fed into a machine learning framework to build IQA model. The model is developed based on LUPI paradigm (ε–SVR+) which utilizes RR measures and privileged information. In testing stage, the RR measures extracted from test images are mapped to quality scores using the constructed SVR+ model.

Feature extraction is an important step to build an IQA model. The features should effectively express the degree and type of distortion independent of the image content. We propose two different feature extraction methods for the predicted and disorderly parts seen the different types of information in these two parts.

The feature extraction module is depicted in Fig. 3. The predicted parts of the reference and distorted images (I_P and $I_P^{'}$) are subjected to a shearlet transform and the mean values of the shearlet coefficients are yielding the quality features (f_P and $f_P^{'}$). The disorderly parts of the reference and distorted images (I_D and $I_D^{'}$) contain uncertainty information of which the Rényi directional entropy is computed to obtain the disorderly features (denoted as f_D and $f_D^{'}$). Next, the difference values between the features are computed in each part to obtain the corresponding quality measures (q_p, q_D). The feature extraction procedures for the two decomposed parts are elaborated in the following subsections.

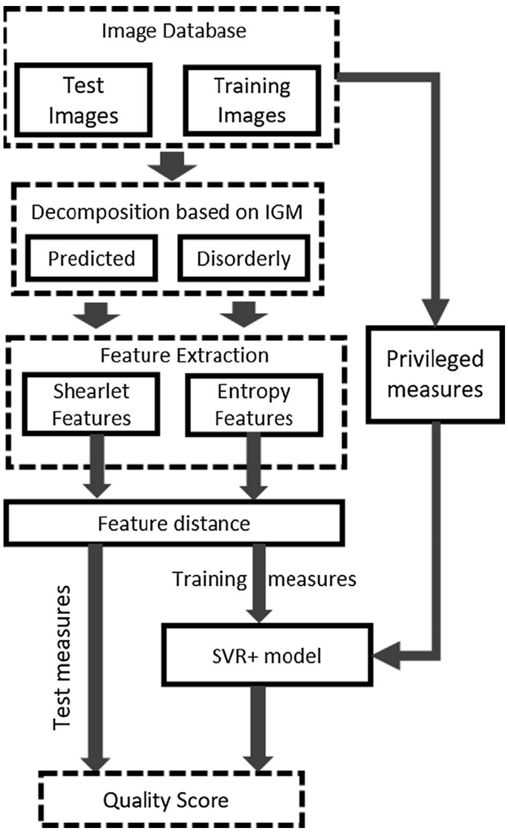

Fig. 2. Overview of the proposed method.

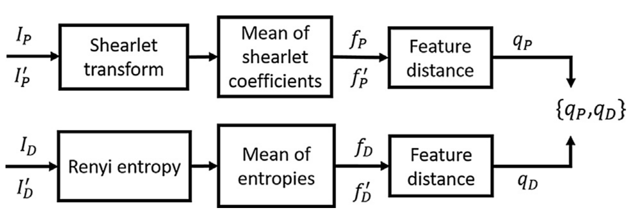

Fig. 3. The feature extraction framework. Inputs: I_P and $I_p^{'}$ are the predicted parts of the reference and distorted images, respectively. I_D and $I_D^{'}$ denote the disorderly portions. Outputs: q_P, q_D are the predicted and disorderly measures.

3.1. Predicted-part features

Here we demonstrate the shearlet transform utilized to extract structural features from the predicted parts of images.

Wavelets can provide a sparse representation of the directional features [39]. To overcome the inherent limitations of wavelets in dealing with multivariate directional data, more advanced, directionality sensitive bases such as curvelets [40], contourlets [41] and shearlets [42] have been developed. Curvelets are introduced as a pyramid of functions defined at multiple scales, locations and orientations in which the rotation is used to capture the directional data. Contourlets are a discrete version of curvelets based on a tree-structured filter bank implementation and they offer efficient numerical methods like standard wavelets do. Shearlets are known as a truly multivariate version of the wavelet framework providing an optimally sparse multi-scale representation of prominent structures such as edges. Compared to curvelets, shearlets have a single or finite set of generation functions. Curvelets involve rotations which cannot preserve the digital lattice. However, shearlets can preserve the structure by parameterization of the directions based on the slope rather than the angles. The contourlet transform utilizes directional filtering to exploit directional data while the shearlet transform takes advantage of the shear matrix, which gives higher flexibility in choosing the number of directions.

The primary visual information of an image is presented in the predicted parts. Such structural information is well-presented in the shearlet domain. The deviations between the shearlet coefficients of I_P and $I_p^{'}$ represent the structural changes caused by the introduced distortions and hence, are also indicative for quality degradation.

Shearlets form an affine system parameterized by three parameters: scale, shear, and translation. The shearlet transform of an image I is defined as:

$$I \rightarrow SH_{\psi}I(b,s,t) = \langle I, \psi_{b,s,t} \rangle \tag{2}$$

where $b > 0$ is the scale parameter, $s \in \mathbb{R}$ denotes the shear parameter and $t \in \mathbb{R}^2$ is the translation parameter. The shearlet coefficients ($\psi_{b,s,t}$) are given by:

$$\begin{gathered}\psi_{b,s,t}(x) = \left|\det L_{b,s}\right|^{-\frac{1}{2}}\psi\left(L_{b,s}^{-1}(x-t)\right) \;\; \textit{where} \\ L_{b,s} = S_s B_b = \begin{bmatrix} b & s\sqrt{b} \\ 0 & \sqrt{b} \end{bmatrix} \\ B_b = \begin{bmatrix} b & 0 \\ 0 & \sqrt{b} \end{bmatrix} \qquad S_s = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix}\end{gathered} \tag{3}$$

$\psi(.)$ is the Meyer wavelet function. To achieve optimal sparsity, the anisotropic dilation matrix B_b enables the multi-scale property and the shear matrix S_s detects directions.

The predicted part of the reference and distorted images are transformed into ten high-pass directional subbands and one low-pass subband by performing a 1-level shearlet transform. The predicted part mainly contains primary high frequency components – such as edges – as well as the low frequency components. Since the distortion impact is more dominant on the high spatial frequency components of an image than on the low frequencies, we used ten directional subbands of the finest – i.e. highest frequency – scale to extract features.

Since distortion can change the energy spectrum of images, the energy of each subband is considered as a feature. The norm-1 energy $e_1(.)$, which is the mean of shearlet coefficients in one subband, is computed as:

$$e_1(k) = \frac{1}{N}\sum_{j=1}^{N}\left|\psi_j^k\right| \tag{4}$$

where ψ_j^k denotes the j-th shearlet coefficient of the k-th subband.

Finally, the normalized difference q_P between the energy values of the k-th subband of the distorted and reference images, respectively $e_1(.)$ and $e_1^{'}(.)$, is calculated as:

$$q_P(k) = \frac{e_1^{'}(k) - e_1(k)}{e_1(k)} \tag{5}$$

Considering a decomposition in 10 subbands, we obtain a subset of 10 measures $\{q_P(1), q_P(2), \cdots, q_P(10)\}$ for the predicted part.

3.2. Disorderly-part features

This section introduces the Rényi directional entropy [31] employed for feature extraction from the disorderly part. Unlike the predicted part distortion, which mostly affects the main structures and image understanding, the distortion of the disorderly portion incurs uncomfortable perception and it distresses our attention. The disorderly portion of an image contains residual uncertainty information and the pixel values in the disorderly image represent the degree of uncertainty. As discussed in Section 2, each distortion has a distinct impact on the amount and composition of information in the disorderly part. In Blur-Noise distorted images, the disorderly part mostly represents the effect of noise by increasing the general entropy in all directions while the Blur-JPEG distortion with blocking artefacts yields heterogeneous changes of information in different directions. Here, the generalized Rényi entropy has been utilized to capture the amount of information changes in various directions.

Entropy is an effective measure of the amount of information in an image and distortion can alter image entropy. Researches have shown a close relationship between image entropy and perceived image quality. Sheikh et al. [43] employed entropy to develop an FR IQA model. Liu et al. [44] extract spatial and spectral entropy features of images for NR IQA.

Gabarda et al. [31] proposed a generalized Rényi entropy approach to achieve directional selectivity. The method extracts the spatial-frequency information of a given 2D image by associating the pixel-level data with a spatial/spatial-frequency distribution called 1-D pseudo-Wigner distribution (PWD) [45]. Each pixel of a 2D image is associated to a vector of a 1-D PWD, which can be set in any direction over the image. Then, the PWD is approximated and the Rényi entropy is measured for every pixel.

The general Rényi entropy (RE) is defined as:

$$RE_{\gamma} = \frac{1}{1-\beta} log_2 \left(\sum_n \sum_k Y^{\beta}[n,k] \right) \tag{6}$$

where Y[n,k] denotes a discrete space-frequency distribution of image, and n, k are space and frequency variables, respectively. Here, $\beta \geq 2$ values are utilized for space–frequency distribution measures. We select $\beta = 3$ for the proposed method.

As mentioned earlier, the PWD is selected to model the discrete space-frequency distribution of the image. A discrete approximation of the PWD can be expressed as follows:

$$W_h[n,k] = 2 \sum_{c=-M/2}^{\frac{M}{2}-1} h[n+c]h^*[n-c]e^{-2i(2\pi c/M)k} \tag{7}$$

where c is the spatial shift, h[n] denotes a vector of gray values of M pixels in a selected direction, and the complex conjugate of h[n] is indicated as $h^*[n]$.

The PWD is computed in a 1-D-oriented window to obtain entropy in a selected direction. $W_h[n,k]$ in (7) represents the distribution in a limited PWD window (spatial interval $(-M/2, M/2+1)$). By moving the directional PWD window over all desired positions in the image, we compute the overall PWD. The directional information can be obtained by rotating the window in various directions. Finally, $W_h[n,k]$ is normalized [31] and associated with $Y[n,k]$ to compute a pixel-wise Rényi entropy.

We measure the pixel-wise Rényi entropy of the disorderly parts in six equally-spaced directions (0°, 30°, 60°, 90°, 120°, and 150°). The mean of the entropy values in each direction j is obtained as disorderly features for the reference and distorted images. Finally, the normalized difference between the features of the distorted and reference image are computed to obtain a quality measure subset q_D:

$$q_D(j) = \frac{\mu'_D(j) - \mu_D(j)}{\mu_D(j)} \tag{8}$$

where $\mu'_D(j)$ and $\mu_D(j)$ denote the mean values of pixel entropies in direction j for the disorderly part of the distorted and reference images, respectively. The feature differences are computed in all directions yielding a set of six measures $\{q_D(1), q_D(2), \cdots, q_D(6)\}$.

4. Quality prediction

We obtained as such ten quality measures from the predicted part $q_P$and six measures from the disorderly part q_D. In total, a subset of 16 measures is provided (q_P, q_D), which is called the standard RR feature set. Since the effects of different distortion types are not similar in the predicted and the disorderly parts, the performance of the sixteen measures will differ depending on the nature of the distortions the image was subjected to. Here, ε–SVR+ (LUPI paradigm) framework [32] is employed to determine an appropriate feature pooling. For quality prediction, we used LUPI to learn a regression model from both standard training data and privileged information. The privileged information is available only at training time and utilized to reduce the error measure of the prediction.

Given training data $\{(x_i, y_i)\}, x_i \in X, y_i \in (0,1)$, the standard ε–SVR aims to find a hyper-plane $f(x) = w.x_i + b$ that has at most ε deviation from the obtained targets y_i(MOS) for the training set and is as flat as possible. This means that as long as the errors are less than ε, they are not taken into consideration. However, any deviation larger than this will not be accepted [46]. The goal of LUPI, which is implemented by the ε–SVR+ algorithm, is to utilize some additional information (i.e. privileged features) during the training phase to further constrain the solution in the feature space X and to build a more sophisticated model. In the situation with privileged information $\widehat{x} \in \widehat{X}$, the triplets $\{(x_i, \widehat{x}_i, y_i)\}, 1 \leq i \leq m$ are given, and three sets of functions are considered.

The first function set lies in the standard space $(w.x_i + b)$, which approximates the decision function. The other two sets are the correction functions to approximate the slack variable functions $\xi_i(\widehat{w}_1, \widehat{b}_1) = \widehat{w}_1.\widehat{x}_i + \widehat{b}_1$ and $\xi_i^*(\widehat{w}_2, \widehat{b}_2) = \widehat{w}_2.\widehat{x}_i + \widehat{b}_2$. Compared to standard SVR, here the slack variables are the correcting functions restricted by privileged information [32]. The optimization problem is formulated as:

$$\begin{aligned} \underset{w, \widehat{w}_1, \widehat{w}_2, b, \widehat{b}_1, \widehat{b}_2}{minimize} \ & \frac{1}{2}\left(||w||^2 + \gamma\left(||\widehat{w}_1||^2 + ||\widehat{w}_2||^2\right)\right) \\ & + C\sum_{i=1}^m \left[\widehat{w}_1.\widehat{x}_i + \widehat{b}_1\right] + C\sum_{i=1}^m \left[\widehat{w}_2.\widehat{x}_i + \widehat{b}_2\right] \end{aligned} \tag{9}$$

$$\begin{aligned} subject\ \ to: \ & y_i - (w, x_i) - b \leq \varepsilon + \widehat{w}_1.\widehat{x}_i + \widehat{b}_1 \\ & (w, x_i) + b - y_i \leq \varepsilon + \widehat{w}_2.\widehat{x}_i + \widehat{b}_2 \\ & \widehat{w}_1.\widehat{x}_i + \widehat{b}_1 \geq 0 \\ & \widehat{w}_2.\widehat{x}_i + \widehat{b}_2 \geq 0 \\ & i = 1 \cdots m \end{aligned}$$

where w represents the weight vector, b is the bias parameter and C denotes the penalty parameter

As mentioned earlier, additional privileged information $\widehat{x}$ is provided in case of SVR+, giving a set of training data $\{(x_i, \widehat{x}_i, y_i)\}$. Here, we utilized FR IQA measures as privileged information. Five FR IQA metrics (FSIM [4], VIF [43], IGM [25], HDR-VDP [47], and GMSD [5]) with high correlation to MOS are selected. Table 1 lists the selected metrics and their correlation behaviour on the MDID

Table 1
Performance of five FR IQA methods on all images of the MDID2015 dataset in terms of PLCC, SROCC, and RMSE.

	PLCC	SROCC	RMSE
FSIM	0.926	0.933	0.125
VIF	0.936	0.937	0.112
IGM	0.876	0.864	0.158
HDR-VDP	0.872	0.859	0.166
GMSD	0.890	0.879	0.149

dataset in terms of Pearson Linear Correlation Coefficients (PLCC), Spearman Rank Order Correlation Coefficients (SROCC), and Root Mean Square Error (RMSE). The FR IQA metrics use full information of the reference image to predict the quality degradation and the state-of-art FR metrics often show better correlation to subjective scores when compared to RR and NR IQA methods. Although, full access to the reference image is not possible in the test stage of the RR scenario, the FR features can be used as additional information during training to build a more sophisticated model. Therefore, in addition to sixteen RR measures (standard data) obtained from the extracted features, five FR measures (privileged data) are also used yielding a 5-D privileged vector (q_{Pr}). The privileged information is only available during the training session and the testing phase is performed by using the trained model and the RR standard measures:

$$Q = \mathbb{P}\left((q_S)_{test}, \mathbb{M}\left[(q_S)_{train}, q_{Pr}\right]\right) \tag{10}$$

where $\mathbb{M}[.]$ is the trained model obtained from the standard $(q_S)_{train}$ and privileged data q_{Pr}, and $\mathbb{P}(.)$ is the prediction function to achieve the quality score Q.

5. Experimental results

In this section, first we demonstrate the efficacy of the feature extraction approach to capture the effect of distortion on the predicted and disorderly portions. Next, the proposed approach is testified on multiply-distorted image datasets.

The performance of the proposed method highly depends on the effectiveness of the features extracted from the images. Various distortion types can change the information in different ways and such changes should be well-represented by the features.

Each multiply-distortion type originates from several single-distortion types possibly interacting with each other when applied to an image, and it is not easy to measure the effect of single distortions, separately. The image decomposition based on IGM is used to better capture the impact of each distortion on the structural and uncertainty information. By using sophisticated feature extraction methods for the two decomposed parts, the multiply-distortion effect can be interpreted and modeled more efficiently.

In the proposed method, the distortion effects in predicted and disorderly parts are presented using shearlet- and entropy-based methods, respectively. Fig. 4 illustrates feature values of the *scrimmage* image from the MLIVE dataset [18] and its five distorted versions (Gaussian Blur (GB), Blur-JPEG, Blur-Noise, JPEG, and Additive White Gaussian Noise (AWGN)). To better visualize the feature changes of distorted images with respect to original-image features, the jth normalized feature response $\widehat{f}$ is computed by dividing each feature value f with the feature value of the original-image f_{org} expressed as: $\widehat{f}(j) = f(j)/f_{org}(j)$. The feature changes on the predicted and disorderly parts are presented under various single and multiply-distortion types. The corresponding 16D feature vector is displayed in which the first 10 features are extracted from the predicted part (shearlet subband features) and the 6 remaining features are derived from the disorderly part (directional entropy features).

Comparing the feature changes due to different distortions, it can be seen that the behaviour of the distortion types differs from each other. AWGN distortion degrades the quality by producing unwanted random data in image. The added information cannot be interpreted by the HVS and causes perceptual quality loss. Based on IGM, AWGN mostly affects the uncertainty information in disorderly part. As shown in Fig. 4, this noise has small impact on the energy of the shearlet information in the predicted part when compared to the original (reference) image while it clearly changes the entropy features in disorderly part. On the other hand, blur distortion removes structural and texture information from predicted and disorderly parts; therefore, the energy in the shearlet subbands and the entropy in the disorderly part is decreased. JPEG distortion mainly causes blocking artifacts in the image with moderate effect on structural and uncertainty information.

As presented by features, each single distortion has its specific effect on the decomposed portions. The distinct behaviour of single distortions is helpful to interpret the multiply-distortion types. For the Blur-Noise multiply-distortion type, we expect to observe the effect of both single distortion types of AWGN and GB together

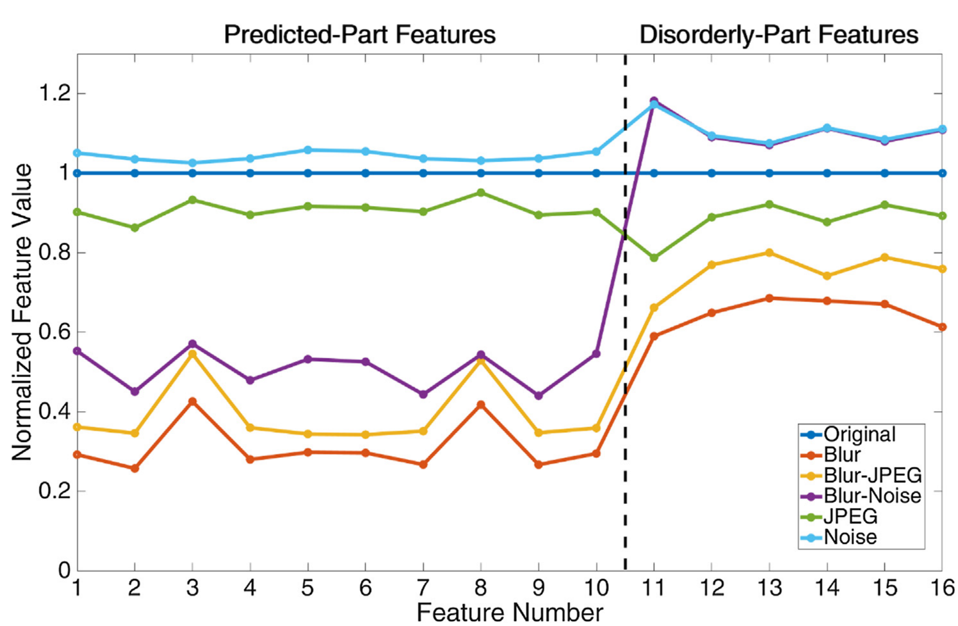

Fig. 4. Feature changes under five distortion types (GB, JPEG, AWGN, Blur-JPEG, and Blur-Noise) for the *scrimmage* image of the MLIVE dataset. Features 1–10 show structural changes in the predicted part and features 11–16 are extracted from the disorderly portion and represent changes in uncertainty information.

with some interactions between them. Subsequently, in the Blur-Noise distorted image of Fig. 4, the blur effect decreases the energy of the shearlet information; however, due to added energy of noise, the amount it decreases is smaller than when compared to a single blur distortion type. On the other hand, in the disorderly portion, noise increases the Rényi entropy to the same extent as the single noise distortion. Comparing the effect of noise and blur in this type of multiply-distortion, the blur effect is more evident in the predicted part as it decreases the subbands energy while entropy values in disorderly part are significantly increased due to the effect of noise distortion. The Blur-JPEG multiply-distortion decreases the feature values in both portions while the values fall in between the feature values of single JPEG and GB distortions.

The impact of different distortions on the composition of the uncertainty information is also well-presented by the extracted features (number 11 to 16 in Fig. 4). The entropy values of disorderly parts are measured in six different directions (0°, 30°, 60°, 90°, 120°, and 150°). Each distortion type has a specific effect on entropy values when compared to the original image. A distortion such as JPEG, with a block-wise effect on the disorderly part, decreases the entropy though the entropy reduction is not equal in all directions.

As it is described, multiply-distortion type images inherit the characteristics of the individual distortions. However, they cannot be classified as one of the single distortion types since they have specific degradation features due to the interaction of the single distortions. As indicated in Fig. 4, the proposed feature extraction method can effectively discriminate between the visual changes of different multiply-distortion and single distortion types.

The performance of the proposed approach is evaluated on multiply-distortion image datasets MDID2015 [19] and MLIVE [18]. The IQA model is trained on MDID2015 dataset, which includes 20 pristine images corrupted by five distortion types: AWGN, GB, Contrast Change (CC), JPEG, and JPEG 2000. Four degradation levels – from slight to severe – are considered for each distortion type. We exclude the images subjected to the CC distortion type and used a total number of 558 distorted images in the experiment. As subjective quality rating, a Differential Mean Opinion Scores (DMOS) is reported for each image.

We randomly partitioned the MDID dataset into train and test subsets; 80% of images is issued for training and the remaining 20% for testing. To ensure the validity of experiment, the train and test images were separated by content and the random partitioning was performed 500 times. From the training set, the IQA models were constructed for application onto the test images. Finally, we report the mean performance indices across 500 experiments.

A mapping function is adopted between objective and subjective scores. Here, we used a 5-parameter logistic function:

$$f(s) = \gamma_1 . \left(\frac{1}{2} - \frac{1}{1 + \exp(\gamma_2 .(s - \gamma_3))} \right) + \gamma_4 . s + \gamma_5 \qquad (11)$$

where $f(s)$ is the mapped score and γ_1 to γ_5 are fitting parameters. Then, several performance evaluation criteria were computed defined as follows. The evaluation indices are defined as follows:

- Pearson Linear Correlation Coefficients (PLCC) denotes the strength of relationship between two sets and defined as:

$$PLCC = \frac{\Sigma_{i=1}^{N}(Q_i - \mu_Q)(Y_i - \mu_Y)}{\sqrt{\Sigma_{i=1}^{N}(Q_i - \mu_Q)^2 \Sigma_{i=1}^{N}(Y_i - \mu_Y)^2}} \qquad (12)$$

where Q_i and Y_i are the objective and subjective quality scores of the ith image and μ denotes average score.

- Root Mean Square Error (RMSE) measures the extent of error between two sets expressed as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Q_i - Y_i)^2}{N}} \qquad (13)$$

- Spearman Rank Order Correlation Coefficients (SROCC) is a rank-order based correlation metric:

$$SROCC = 1 - \frac{6\sum_{i=1}^{N}\delta_i^2}{N\left(N^2 - 1\right)} \qquad (14)$$

where δ_i is the difference between ranks of ith image $[rank(Q_i) - rank(Y_i)]$ in objective and subjective score sets.

PLCC and RMSE indicate the accuracy of prediction and SROCC evaluates the prediction monotonicity. We also report Standard Deviation (SD) of PLCC values across 500 trails. The smaller SD means better stability of the model to variations in training data.

The performance of the proposed method is compared with several FR, RR and NR IQA metrics. The competing FR metrics are Feature SIMilarity (FSIM) [4], Visual Information Fidelity (VIF) [43], High Dynamic Range Visual Difference Predictor (HDR-VDP 2.2) [47], Internal Generative Mechanism model (IGM) [25], Peak Signal-to-Noise Ratio (PSNR), and Structural SIMilarity index (SSIM) [48]. The RR approaches include Reduced Reference Wavelet Domain based quality (RRWD) [49], Reduced Reference Entropic Differencing (RRED) [7] and Orientation Selectivity Based Visual Pattern (OSVP) [14]. The competing NR metrics include two multiply-distortion methods: Five-Step Blind Metric (FISBLIM) [15] and Six-Step Blind Metric (SISBLIM) [16] and four general-purpose metrics: NR Free Energy based Robust Metric (NFERM), High Order Statistics Aggregation (HOSA) [50], Multi-task End-to-end Optimized deep neural Network (MEON) [51], Distortion Identification-based Image Verity and INtegrity Evaluation (DIIVINE) [8], Spatial-Spectral Entropy-based Quality (SSEQ) [44], Natural Image Quality Evaluator (NIQE) [52], and Blind/Referenceless Image Spatial QUality Evaluator (BRISQUE) [53]. In addition to the target models (SVR and SVR+), we also trained our designed features using Extreme Learning Machine (ELM) framework to better validate the effectiveness of the designed features and illustrate the advantage of LUPI paradigm compared to the conventional learning methods. ELM is developed as a class of single-hidden layer feedforward neural network and it offers a good generalization ability with high training speed.

Table 2 summarizes the performance of the competing metrics. The IQA model learned by the LUPI paradigm (SVR+) improves the performance when compared to the standard SVR and ELM. The $RRSEA_{SVR+}$ model with a PLCC of 0.843 and SROCC of 0.831 outperforms $RRSEA_{SVR}$ and $RRSEA_{ELM}$. In contrast, $RRSEA_{SVR+}$ has smaller RMSE, which means less prediction error. The standard deviation of the PLCC indices is also decreased by using ε–SVR+; implying a more stable model and less dependency on the training set. Comparing SVR- and ELM-based metrics, the SVR model performs slightly better. As shown in the table, the proposed method significantly outperforms PSNR and SSIM FR metrics. VIF and FSIM show the highest performance among all metrics and the performance of $RRSEA_{SVR+}$ is comparable with the HDR-VDP metric.

The proposed method delivers much better results compared to all other NR approaches. Our method outperforms the competing RR IQA approaches. Compared to RRED, as one of the best available RR metrics, the $RRSEA_{SVR+}$ shows higher correlation and much lower RMSE. In addition, lower SD values of the proposed method compared to RRED indicate lower dependency of the proposed model to the training data.

Table 2
Performance of different IQA metrics on MDID2015. Mean PLCC, SROCC, and RMSE as well as the standard deviation of PLCC values (SD) are computed across 500 train-test trials.

IQA Metrics	Type	PLCC	SROCC	RMSE	SD
FSIM	FR	0.932	0.926	0.160	0.018
VIF	FR	0.940	0.937	0.148	0.025
HDR-VDP	FR	0.862	0.866	0.190	0.044
IGM	FR	0.873	0.884	0.171	0.045
SSIM	FR	0.627	0.643	0.326	0.127
PSNR	FR	0.644	0.660	0.309	0.144
$RRSEA_{ELM}$	RR	0.791	0.794	0.255	0.061
$RRSEA_{SVR}$	RR	0.815	0.807	0.209	0.052
$RRSEA_{SVR+}$	RR	0.843	0.831	0.185	0.047
RRWD	RR	0.672	0.682	0.344	0.073
RRED	RR	0.786	0.834	0.391	0.158
OSVP	RR	0.685	0.713	0.293	0.081
FISBLIM	NR	0.602	0.618	0.313	0.059
SISBLIM	NR	0.661	0.679	0.325	0.057
NFERM	NR	0.579	0.541	0.332	0.062
HOSA	NR	0.687	0.662	0.323	0.059
MEON	NR	0.610	0.554	0.354	0.058
SSEQ	NR	0.571	0.558	0.343	0.056
NIQE	NR	0.670	0.664	0.357	0.060
DIIVINE	NR	0.588	0.585	0.306	0.059
BRISQUE	NR	0.590	0.535	0.324	0.061

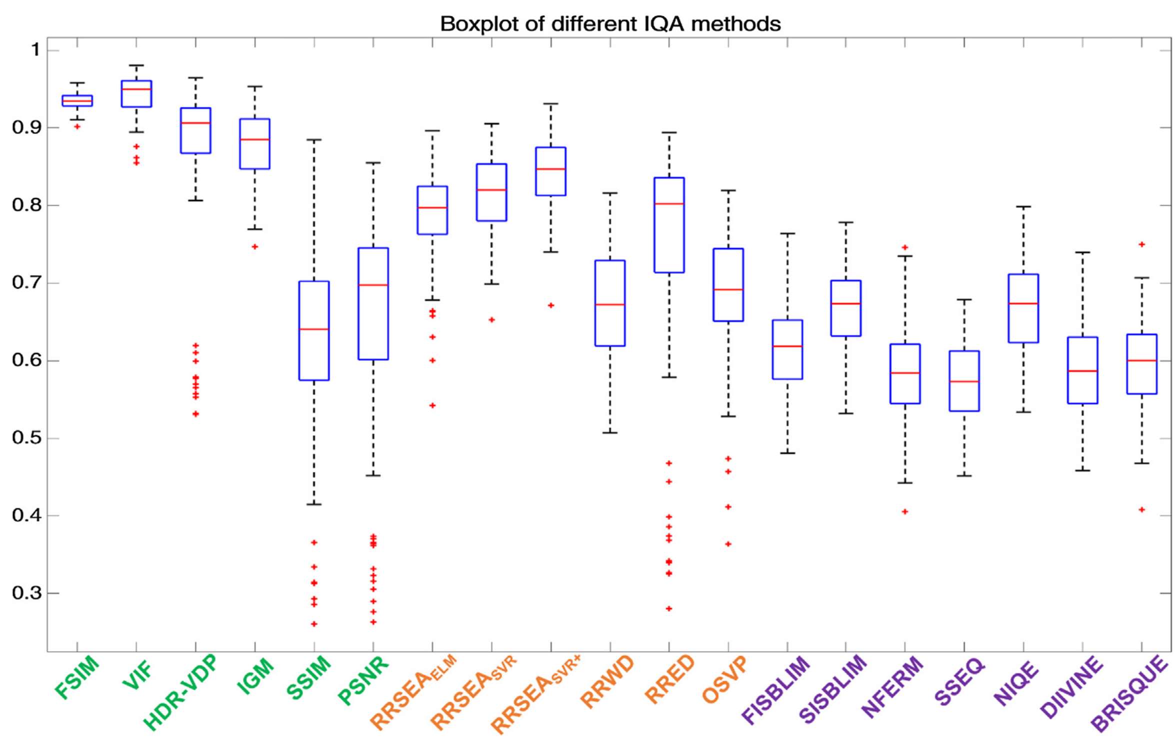

Fig. 5. Boxplot comparing the PLCC distribution of the competing algorithms over 500 train-test trials on the MDID2015 dataset.

The distribution of PLCC values for different methods is visualized using boxplots in Fig. 5. Apart from the mean correlation over 500 trials (shown in Table 2), the boxplots show the median (red lines), the standard error, and the maximum/minimum of 500 PLCC correlations. The outliers are specified by red '+' symbols. Comparing VIF and FSIM, as the best metrics, VIF has higher median PLCC while FSIM depicts a lower error and standard deviation. The proposed approach outperforms all NR metrics. The boxplot representation of RRED metrics implies its larger prediction error compared to the proposed model.

To provide a clear understanding on the performance comparison and statistical significance of competing approaches, a two-sample student *t*-test is conducted [54]. The *t*-test is computed based on the SROCC values between predicted scores and subjective DMOS over 500 train-test trials. The statistical performance is listed in Table 3 in which value '1' implies that an IQA metric on the horizontal axis is statistically better than a method on the vertical axis. The symbol '0' is used when the statistical significance is indistinguishable (no statistically significant distance) and symbol '−1' means an IQA metric on the horizontal axis is worse than a metric on the vertical axis. We set the confidence level to 95% for all comparisons.

As it is observed in Table 3, the proposed $RRSEA_{SVR+}$ method is statistically better than RRED, OSVP and RRWD. The $RRSEA_{SVR+}$ model is also statistically better than all NR approaches. In comparison with FR metrics, the $RRSEA_{SVR+}$ can perform as good as the

Table 3
Comparison of statistical significance using two-sample student *t*-test. A value '1' means the IQA method in row is statistically better than the method in column; '0' indicates no significant difference between a row and a column algorithm. '−1' denotes the algorithm in row is inferior to the column.

	FSIM	VIF	HDR-VDP	IGM	SSIM	PSNR	$RRSEA_{ELM}$	$RRSEA_{SVR}$	$RRSEA_{SVR+}$	RRWD	RRED	OSVP	FISBLIM	SISBLIM	NFERM	SSEQ	NIQE	DIIVINE	BRISQUE
FSIM	0	-1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
VIF	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
HDR-VDP	-1	-1	0	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
IGM	-1	-1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
SSIM	-1	-1	-1	-1	0	0	-1	-1	-1	-1	-1	-1	0	-1	1	1	-1	1	1
PSNR	-1	-1	-1	-1	0	0	-1	-1	-1	0	-1	-1	1	0	1	1	0	1	1
$RRSEA_{ELM}$	-1	-1	-1	-1	1	1	0	0	-1	1	0	1	1	1	1	1	1	1	1
$RRSEA_{SVR}$	-1	-1	-1	-1	1	1	0	0	-1	1	0	1	1	1	1	1	1	1	1
$RRSEA_{SVR+}$	-1	-1	0	-1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1
RRWD	-1	-1	-1	-1	1	0	-1	-1	-1	0	-1	0	1	0	1	1	0	1	1
RRED	-1	-1	-1	-1	1	1	0	0	-1	1	0	1	1	1	1	1	1	1	1
OSVP	-1	-1	-1	-1	1	1	-1	-1	-1	0	-1	0	1	1	1	1	0	1	1
FISBLIM	-1	-1	-1	-1	0	-1	-1	-1	-1	-1	-1	-1	0	-1	1	1	-1	1	1
SISBLIM	-1	-1	-1	-1	1	0	-1	-1	-1	0	-1	-1	1	0	1	1	0	1	1
NFERM	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	-1	0	0
SSEQ	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	-1	0	0
NIQE	-1	-1	-1	-1	1	0	-1	-1	-1	0	-1	0	1	0	1	1	0	1	1
DIIVINE	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	-1	0	0
BRISQUE	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	-1	0	0

state-of-art HDR-VDP metric and outperforms PSNR and SSIM. VIF is statistically superior to all competing methods. Among NR approaches, the NIQE is better than other NR IQA algorithms and it is statistically close to OSVP, RRWD, and PSNR. The RRSEA models trained by SVR and ELM are comparable and there is no significant statistical difference between these two model. The results confirm that by using additional priviledged features in SVR+ framework, we can achieve higher performance than traditional SVR and ELM learning approaches.

To evaluate the dependency of the proposed model on the number of training images, other partitioning strategies have also been taken into account. In addition to the 80–20% train-test split, we considered 70–30%, 60–40%, 50–50%, 40–60%, and 30–70% splitting and report on the amount of performance change.

Table 4 reports the PLCC and SROCC values obtained by changing the train-test portions. It is shown that the performance gradually drops by decreasing the number of training images. We also compute the amount of performance loss compared to 80–20% correlation for each evaluation index:

$$Loss\% = \frac{Corr_{x\%-y\%} - Corr_{80\%-20\%}}{Corr_{80\%-20\%}} \times 100 \qquad (15)$$

where $Corr_{x\%\text{-}y\%}$ denotes the correlation over 500 train-test by choosing *x*% of images for training and *y*% for testing. From Table 4, it can be observed that the performance drops much faster for RRED. Also, the proposed models perform significantly better than RRED when the number of training images are decreased. Comparing the two $RRSEA_{SVR}$ and $RRSEA_{SVR+}$ models for all splitting methods, it is observed that the proposed model based on SVR+ depicts a slower performance drop than the standard SVR model. For example, in 50%-50% splitting, the amount of performance loss of $RRSEA_{SVR}$ in terms of PLCC and SROCC is respectively, 4.9% and 4.3% while respectively, 3.4% PLCC and 3.0% SROCC loss are reported for the $RRSEA_{SVR+}$ model. The results illustrate that both models can still perform reasonably well when the amount of training data is decreasing. The trained model based on SVR+ is more robust to training image changes than the standard SVR model.

Table 4
Performance of the RRED, $RRSEA_{SVR}$, and $RRSEA_{SVR+}$ approaches considering various train%-test% partitioning strategies. The performance loss compared to 80%-20% method is also reported.

	PLCC	Loss %	SROCC	Loss %
70% - 30%				
$RRSEA_{SVR}$	0.804	1.3	0.798	1.1
$RRSEA_{SVR+}$	0.834	1.1	0.825	0.7
RRED	0.768	2.3	0.809	2.9
60% - 40%				
$RRSEA_{SVR}$	0.790	3.0	0.783	2.9
$RRSEA_{SVR+}$	0.821	2.6	0.811	2.4
RRED	0.740	5.8	0.779	6.5
50% - 50%				
$RRSEA_{SVR}$	0.775	4.9	0.772	4.3
$RRSEA_{SVR+}$	0.814	3.4	0.806	3.0
RRED	0.715	9.0	0.745	10.6
40% - 60%				
$RRSEA_{SVR}$	0.765	6.1	0.759	5.9
$RRSEA_{SVR+}$	0.798	5.3	0.789	5.0
RRED	0.658	16.2	0.692	17.0
30% - 70%				
$RRSEA_{SVR}$	0.747	8.3	0.741	8.2
$RRSEA_{SVR+}$	0.782	7.2	0.781	6.0
RRED	0.641	18.4	0.682	18.2

IQA models are expected to be dataset-independent. It means that a model trained on one dataset should not be specific to the contents of that dataset and it needs to perform well also on other datasets. To demonstrate that the trained model is generic, the parameters trained on the MDID dataset were tested on the MLIVE dataset. The MLIVE consists of 15 reference images subjected to two multiply-distortion types; Blur-JPEG and Blur-Noise. A total number 450 images with assigned subjective ratings are provided. Table 5 reports the performance comparison on MLIVE dataset. The performance indices of the proposed method dropped slightly due to some differences in simulated distortions. However, the correlation results confirm that the overall performance is still consistently high. Fig. 6 indicates the scatter plot of the predicted scores versus the subjective MOS for HDR-VDP, FSIM, SSIM, PSNR, OSVP, and $RRSEA_{SVR+}$ approaches on the MLIVE dataset.

Table 5
Performance comparison of IQA methods on the MLIVE dataset.

IQA metric	Type	PLCC	SROCC	RMSE
FSIM	FR	0.893	0.879	8.59
VIF	FR	0.902	0.888	8.43
HDR-VDP	FR	0.918	0.899	8.41
IGM	FR	0.881	0.853	8.85
SSIM	FR	0.733	0.646	12.83
PSNR	FR	0.739	0.672	12.75
$RRSEA_{SVR}$	RR	0.797	0.788	11.05
$RRSEA_{SVR+}$	RR	0.824	0.811	10.34
RRWD	RR	0.740	0.706	12.36
RRED	RR	0.819	0.791	10.48
OSVP	RR	0.775	0.732	11.53
FISBLIM	NR	0.782	0.758	11.50
SISBLIM	NR	0.765	0.743	11.98
NFERM	NR	0.613	0.598	14.41
HOSA	NR	0.812	0.802	10.64
MEON	NR	0.768	0.753	11.66
SSEQ	NR	0.362	0.258	16.98
NIQE	NR	0.747	0.721	12.19
DIIVINE	NR	0.709	0.664	13.22
BRISQUE	NR	0.552	0.512	15.37

The quality metrics designed for multi-distorted images should also perform well on single distortion datasets. As shown earlier in Fig. 4, the proposed features can represent the characteristics of various single and multiple distortion types. Table 6 shows the performance results on LIVE [55] image dataset which consists of 779 distorted images with five types of single distortions – JP2K, JPEG, AWGN, GB, and Fast Fading (FF). The results indicate good performance of the RRSEA on images corrupted by single distortion types.

In communication systems, RR features extracted from reference images are transmitted from sender side to the receiver side where the distorted image is available for quality assessment. Therefore, it is highly desirable to send quantized RR features (with smaller size) without a significant loss of performance. Here, the efficacy of the RRSEA metric is tested when the number of bits per feature (bpf) is changing. An original RR feature is considered to have a size of 32-bit. Fig. 7 presents the bpf vs. SROCC diagram where the number of bits assigned to each feature is changing from 32 to 2 bits. As shown in the figure, the RRSEA model can still perform well when the number of bits is decreasing and there is no significant performance drop by using 8 bits instead of 32 bits per feature. Considering the sixteen 8-bit features extracted from reference image, the proposed RRSEA model only needs 128 bits of data from the sender side for quality assessment.

In Table 7, we summarized the feature extraction methods of different competing RR approaches. In applications such as quality monitoring of visual communications systems, the RR methods are more preferred than FR metrics due to much lower data rate of RR features. The proposed method only needs 16 features from the reference image (at sender side) to predict the quality which is quite reasonable compared to other RR methods.

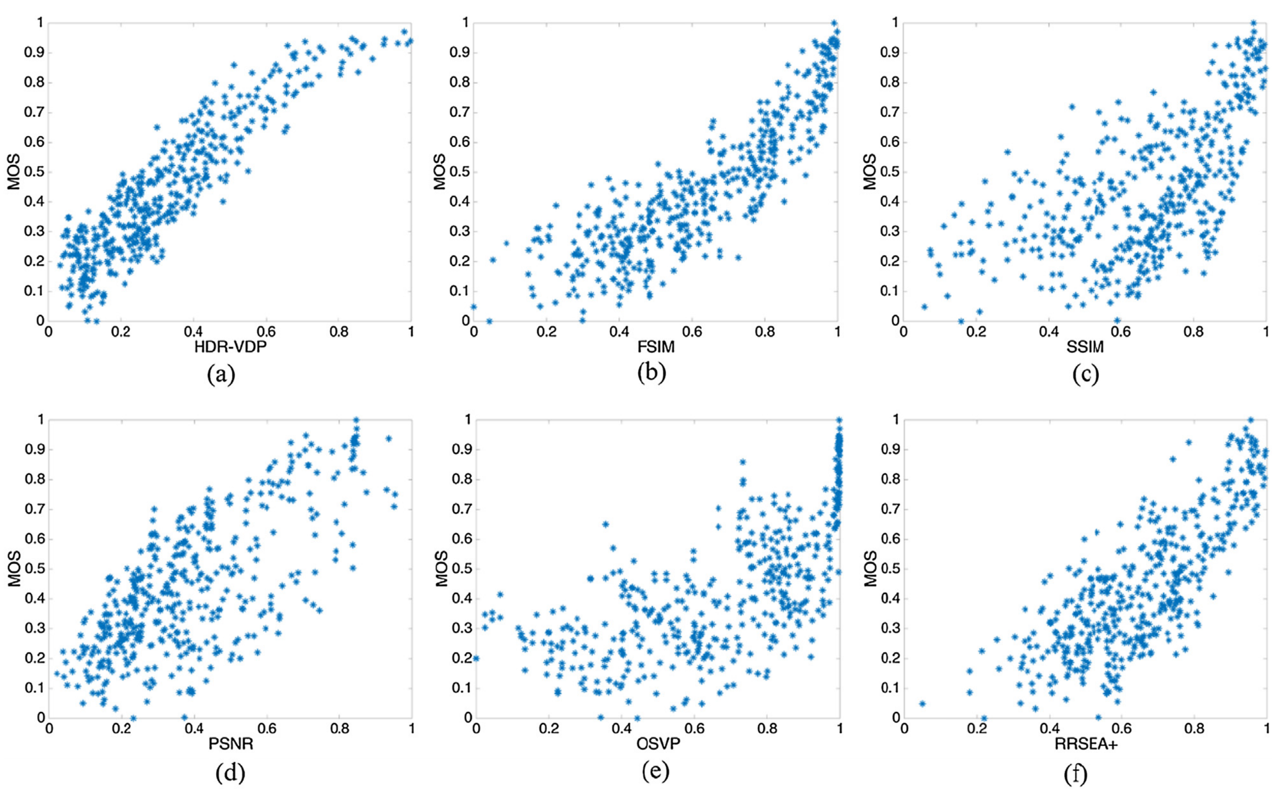

Fig. 6. Scatter plots of (a) HDR-VDP, (b) FSIM, (c) SSIM, (d) PSNR, (e) OSVP, and (f) $RRSEA_{SVR+}$ approaches for 450 images of the MLIVE dataset.

Table 6
Performance evaluation on LIVE single distortion image dataset. Mean PLCC and SROCC are computed across 500 train-test trials.

PLCC	JP2K	JPEG	WN	BLUR	FF	ALL
RRED	0.963	0.979	0.985	0.969	0.923	0.949
BRISQUE	0.923	0.973	0.985	0.950	0.903	0.942
MEON	0.926	0.979	0.986	0.951	0.906	0.946
HOSA	0.938	0.976	0.988	0.979	0.921	0.952
NFERM	0.954	0.971	0.985	0.938	0.887	0.944
DIIVINE	0.923	0.934	0.986	0.937	0.891	0.927
SSEQ	0.943	0.972	0.970	0.935	0.916	0.936
NIQE	0.926	0.952	0.976	0.943	0.879	0.905
RRSEA	0.955	0.986	0.983	0.977	0.912	0.951
SROCC	**JP2K**	**JPEG**	**WN**	**BLUR**	**FF**	**ALL**
RRED	0.958	0.976	0.978	0.963	0.916	0.941
BRISQUE	0.916	0.964	0.978	0.941	0.875	0.939
MEON	0.912	0.967	0.983	0.934	0.891	0.943
HOSA	0.933	0.954	0.972	0.952	0.900	0.950
NFERM	0.941	0.964	0.983	0.921	0.862	0.940
DIIVINE	0.912	0.920	0.981	0.937	0.868	0.925
SSEQ	0.941	0.953	0.975	0.920	0.905	0.933
NIQE	0.918	0.942	0.971	0.933	0.863	0.908
RRSEA	0.945	0.981	0.980	0.969	0.901	0.947

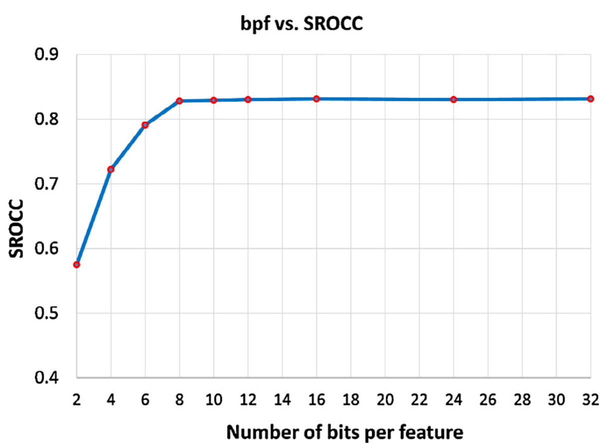

Fig. 7. Performance of the proposed method on MDID2015 dataset for different number of bits (2–32 bits) assigned per feature.

The required computation time of the proposed method is reported in Table 8. The Matlab code was executed for input images of size 512×384, 10 repetitions, and the average processing time is reported. The test was performed on a Windows laptop with 16 GB RAM and a Core i7-2.7 GHz CPU. Although RRSEA is not developed under the constraint of real-time application, the computation time is still reasonable and it is faster than DIIVINE and NFERM metrics. Evidently, the processing time can be improved by optimizing the implementation.

6. Conclusion

In this paper, a novel RR IQA method is proposed to predict the quality of multiply-distorted images. Since multiple distortion types can interact with each other when added to an image, it is more challenging to interpret the degradation effect on multiply-distorted images. To better model the effect of distortions, we decomposed an input image into predicted and disorderly parts based on an AR procedure. The degradation of primary visual information in the predicted part is measured using a shearlet representation and the disorderly-part information is obtained in several directions by deploying Rényi entropy. Finally, we proposed the use of the LUPI paradigm (SVR+) to train the quality prediction model. The SVR+ framework utilizes FR measures as privileged information during training. Experimental results indicate good performance of the proposed algorithm independent from the trained dataset. The proposed model outperforms several FR and state-of-the-art RR and NR metrics on MDID and MLIVE

Table 7
The extracted features of different RR methods.

RRWD	18 features by modeling the distributions of coefficients in 6 wavelet subbands. In each subband, the following features are extracted: • The parameters of Generalized Gaussian Distribution (GGD) model (2 features) • The prediction error computed by Kullback–Leibler Divergence (KLD) (1 feature)
RRED	The entropies of a wavelet subband coefficients are computed as features. • The number of features is equal to size of a selected subband divided by block size (3×3). (number of features are typically around 1/18 of the image size)
OSVP	The visual content is extracted by OSVP analysis and mapped to histogram bins. • 9 histogram bins are utilized as features
RRSEA	16 quality-relevant features obtained by IGM-based image decomposition • 10 features from shearlet subbands of predicted part • 6 directional entropy-based features from disorderly part

Table 8
Comparison of average required computation time (s).

	HDR-VDP	FSIM	OSVP	HOSA	DIIVINE	NFERM	RRSEA
Time (sec)	1.81	0.33	0.14	0.65	12.68	26.75	7.38

multiply-distortion image datasets. The results illustrate that using privileged information during training of SVR+ model can help to construct a more sophisticated model and improve the quality prediction accuracy.

Declarations of interest

None.

Acknowledgements

The research leading to these results has received funding from the European Research Council under the European Unions Seventh Framework Programme (FP7/2007-2013)/ERC Grant Agreement Nr. 617779 (INTERFERE) and Horizon 2020 Research and Innovation Programme under Grant Agreement N688619 (ImmersiaTV).

References

[1] W. Lin, C.C. Jay Kuo, Perceptual visual quality metrics: a survey, J. Vis. Commun. Image Represent. 22 (4) (2011) 297–312.
[2] L. Liu, Y. Hua, Q. Zhao, H. Huang, A.C. Bovik, Blind image quality assessment by relative gradient statistics and adaboosting neural network, Signal Process. Image Commun. 40 (2016) 1–15.
[3] Y. Ding, Y. Zhao, X. Zhao, Image quality assessment based on multi-feature extraction and synthesis with support vector regression, Signal Process. Image Commun. 54 (2017) 81–92.
[4] L. Zhang, L. Zhang, X. Mou, D. Zhang, FSIM: a feature similarity index for image quality assessment, IEEE Trans. Image Process. 20 (8) (2011) 2378–2386.
[5] W. Xue, L. Zhang, X. Mou, A.C. Bovik, Gradient magnitude similarity deviation: a highly efficient perceptual image quality index, IEEE Trans. Image Process. 23 (2) (2014) 668–695.
[6] Y. Zhang, T.D. Phan, D.M. Chandler, Reduced-reference image quality assessment based on distortion families of local perceived sharpness, Signal Process. Image Commun. 55 (2017) 130–145.
[7] R. Soundararajan, A.C. Bovik, RRED indices: reduced reference entropic differencing for image quality assessment, IEEE Trans. Image Process. 21 (2) (2012) 517–526.
[8] A.K. Moorthy, A.C. Bovik, Blind image quality assessment: from natural scene statistics to perceptual quality, IEEE Trans. Image Process. 20 (12) (2011) 3350–3364.
[9] S. Mahmoudpour, M. Kim, No-reference image quality assessment in complex-shearlet domain, Signal, Image Video Process. 10 (8) (2016) 1465–1472.
[10] K. Gu, D. Tao, J.F. Qiao, W. Lin, Learning a no-reference quality assessment model of enhanced images with big data, IEEE Trans. Neural Networks Learn. Syst. 29 (4) (2018) 1301–1313.
[11] K. Gu, J. Zhou, J.F. Qiao, G. Zhai, W. Lin, A.C. Bovik, No-reference quality assessment of screen content pictures, IEEE Trans. Image Process. 26 (8) (2017) 4005–4018.
[12] K. Gu, V. Jakhetiya, J.F. Qiao, X. Li, W. Lin, D. Thalmann, Model-based referenceless quality metric of 3D synthesized images using local image description, IEEE Trans. Image Process. 27 (1) (2018) 394–405.
[13] Z. Wang, G. Wu, H.R. Sheikh, E.P. Simoncelli, E.H. Yang, A.C. Bovik, Quality-aware images, IEEE Trans. Image Process. 15 (6) (2006) 1680–1689.
[14] J. Wu, W. Lin, G. Shi, L. Li, Y. Fang, Orientation selectivity based visual pattern for reduced-reference image quality assessment, Inf. Sci. (Ny) 351 (2016) 18–29.
[15] A. Rehman, Z. Wang, Reduced-reference image quality assessment by structural similarity estimation, IEEE Trans. Image Process. 21 (8) (2012) 3378–3389.
[16] D.M. Chandler, Seven challenges in image quality assessment: past, present, and future research, ISRN Signal Process. 2013 (2013) 1–53.
[17] K. Gu, G. Zhai, X. Yang, W. Zhang, Hybrid no-reference quality metric for singly and multiply distorted images, IEEE Trans. Broadcast. 60 (3) (2014) 555–567.
[18] D. Jayaraman, A. Mittal, A.K. Moorthy, A.C. Bovik, Objective quality assessment of multiply distorted images, in: Conference Record – Asilomar Conference on Signals, Systems and Computers, 2012, pp. 1693–1697.
[19] W. Sun, F. Zhou, Q. Liao, MDID: A multiply distorted image database for image quality assessment, Pattern Recognit. 61 (2017) 153–168.
[20] L. Zhang, L. Zhang, X. Mou, RFSIM: A feature based image quality assessment metric using Riesz transforms, in: Proceedings – International Conference on Image Processing, ICIP, 2010, pp. 321–324.
[21] Q. Li, W. Lin, Y. Fang, No-reference quality assessment for multiply-distorted images in gradient domain, IEEE Signal Process. Lett. 23 (4) (2016) 541–545.
[22] N. Sarafianos, C. Nikou, I.A. Kakadiaris, Predicting privileged information for height estimation, Proc. – Int. Conf. Pattern Recognit. (2017) 3115–3120.
[23] K. Friston, J. Kilner, L. Harrison, A free energy principle for the brain, J. Physiol. Paris 100 (1–3) (2006) 70–87.
[24] K. Friston, The free-energy principle: a rough guide to the brain?, Trends Cogn Sci. 13 (7) (2009) 293–301.
[25] J. Wu, W. Lin, G. Shi, A. Liu, Perceptual quality metric with internal generative mechanism, IEEE Trans. Image Process. 22 (1) (2013) 43–54.
[26] K. Gu, G. Zhai, X. Yang, W. Zhang, Using free energy principle for blind image quality assessment, IEEE Trans. Multimed. 17 (1) (2015) 50–63.
[27] K. Gu, G. Zhai, W. Lin, X. Yang, W. Zhang, No-reference image sharpness assessment in autoregressive parameter space, IEEE Trans. Image Process. 24 (10) (2015) 3218–3231.
[28] Y. Liu, G. Zhai, K. Gu, X. Liu, D. Zhao, W. Gao, Reduced-reference image quality assessment in free-energy principle and sparse representation, IEEE Trans. Multimed. 20 (2) (2018) 379–391.
[29] J. Fu, H. Wang, L. Zuo, Blind image quality assessment for multiply distorted images via convolutional neural networks, 2016 IEEE Int. Conf. Acoust. Speech Signal Process, pp. 1075–1079, 2016.
[30] W.Q. Lim, The discrete shearlet transform: a new directional transform and compactly supported shearlet frames, IEEE Trans. Image Process. 19 (5) (2010) 1166–1180.
[31] S. Gabarda, G. Cristóbal, Blind image quality assessment through anisotropy, J. Opt. Soc. Am. A. Opt. Image Sci. Vis. 24 (12) (2007) B42–B51.
[32] V. Vapnik, A. Vashist, A new learning paradigm: learning using privileged information, Neural Networks 22 (5–6) (2009) 544–557.
[33] I.A. Kakadiaris, N. Sarafianos, C. Nikou, Show me your body: gender classification from still images, in: Proceedings - International Conference on Image Processing, ICIP, 2016, vol. 2016–Augus, pp. 3156–3160.
[34] X. Xu, W. Li, D. Xu, Distance metric learning using privileged information for face verification and person re-identification, IEEE Trans. Neural Networks Learn. Syst. 26 (12) (2015) 3150–3162.
[35] H. Yang, I. Patras, Privileged information-based conditional structured output regression forest for facial point detection, IEEE Trans. Circuits Syst. Video Technol. 25 (9) (2015) 1507–1520.
[36] D.C. Knill, A. Pouget, The Bayesian brain: the role of uncertainty in neural coding and computation, Trends Neurosci. 27 (12) (2004) 712–719.
[37] K. Friston, The free-energy principle: a unified brain theory?, Nat Rev. Neurosci. 11 (2) (2010) 127–138.
[38] J. Wu, W. Lin, Y. Fang, L. Li, G. Shi, S. Issac Niwas, Visual structural degradation based reduced-reference image quality assessment, Signal Process. Image Commun. 47 (2016) 16–27.
[39] M. Wakin, Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity [Book Reviews], vol. 28, no. 5. 2011.
[40] E. Candès, L. Demanet, D. Donoho, L. Ying, Fast discrete curvelet transforms, Multiscale Model. Simul. 5 (3) (2006) 861–899.
[41] M.N. Do, M. Vetterli, The contourlet transform: an efficient directional multiresolution image representation, IEEE Trans. Image Process. 14 (12) (2005) 2091–2106.
[42] G. Kutyniok, W.Q. Lim, X. Zhuang, Digital shearlet transforms, in: Shearlets: Multiscale Analysis for Multivariate Data, 2012, pp. 239–282.
[43] H.R. Sheikh, A.C. Bovik, Image information and visual quality, IEEE Trans. Image Process. 15 (2) (2006) 430–444.
[44] L. Liu, B. Liu, H. Huang, A.C. Bovik, No-reference image quality assessment based on spatial and spectral entropies, Signal Process. Image Commun. 29 (8) (2014) 856–863.
[45] T.A.C.M. Claasen, W.F.G. Mecklenbrauker, The Wigner distribution – a tool for time-frequency signal analysis, Philips J. Res. 35 (6) (1980) 372–389.
[46] C. Chang, C. Lin, LIBSVM: a library for support vector machines, ACM Trans. Intell. Syst. Technol. 2 (2013) 1–39.
[47] M. Narwaria, R.K. Mantiuk, M.P. Da Silva, P. Le Callet, HDR-VDP-2.2: a calibrated method for objective quality prediction of high-dynamic range and standard images, J. Electron. Imaging 24 (1) (2015) 10501.
[48] 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.
[49] Z. Wang, E.P. Simoncelli, Reduced-reference image quality assessment using a wavelet-domain natural image statistic model, PIE 5666 Hum. Vis. Electron. Imaging X 5666 (2005) 149–159.
[50] J. Xu, P. Ye, Q. Li, H. Du, Y. Liu, D. Doermann, Blind image quality assessment based on high order statistics aggregation, IEEE Trans. Image Process. 25 (9) (2016) 4444–4457.
[51] K. Ma, W. Liu, K. Zhang, Z. Duanmu, Z. Wang, W. Zuo, End-to-end blind image quality assessment using deep neural networks, IEEE Trans. Image Process. 27 (3) (2017) 1202–1213.
[52] A. Mittal, R. Soundararajan, A.C. Bovik, Making a 'completely blind' image quality analyzer, IEEE Signal Process. Lett. 20 (3) (2013) 209–212.
[53] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, IEEE Trans. Image Process. 21 (12) (2012) 4695–4708.
[54] D.J. Sheskin, Handbook of parametric and nonparametric statistical procedures, vol. 46, 2004.
[55] H.R. Sheikh, Z. Wang, A.C. Bovik, L.K. Cormack, Image and video quality assessment research at LIVE [Online]. Available: <htp://live.ece.utexas.edu/research/quality/>.