Reduced-Reference Image Quality Assessment based on Internal Generative Mechanism utilizing Shearlets and Rényi Entropy Analysis

Saeed Mahmoudpour
(1) Vrije Universiteit Brussel (VUB),
Department of Electronics and Informatics,
Pleinlaan 2, B-1050 Brussels, Belgium
(2) imec, Kapeldreef 75, B-3001 Leuven, Belgium
Email: smahmoud@etrovub.be

Peter Schelkens
(1) Vrije Universiteit Brussel (VUB),
Department of Electronics and Informatics,
Pleinlaan 2, B-1050 Brussels, Belgium
(2) imec, Kapeldreef 75, B-3001 Leuven, Belgium
Email: pschelke@etrovub.be

***Abstract*—During acquisition, processing, compression and transmission, images may be corrupted by multiple distortions such as blur, noise or compression artefacts. However, most of the existing image quality assessment (IQA) methods are designed for images degraded by a single distortion type. This paper proposes a reduced-reference (RR) IQA method for quality assessment of multiply distorted images. The method extracts a number of quality-characterizing features from the reference and the distorted images for quality prediction. Based on internal generative mechanism (IGM) theory, the images are decomposed first into their predicted and disorderly portions. Next, a number of quality-characterizing features are extracted from each portion and feature differences are computed between the reference and distorted images. Finally, support vector regression (SVR) is adopted to obtain a quality score. Experimental results on public multiply-distorted image databases, namely MDID2015 and MLIVE, show that the proposed method is well-correlated with subjective ratings and outperforms several IQA methods.**

***Keywords*—image quality; reduced-reference; shearlet transform; entropy; support vector regression; internal generative mechanism theory**

I. Introduction

Following to the rapid advances in multimedia technology and wide usage of smart phones, digital images have become a prevalent medium which can be captured, stored and shared easily. Despite of this progress, digital images are subjected to various distortions in end-to-end application chains. Distortions such as blur, noise, blocking and ringing artifacts can seriously affect the perceived image quality. Therefore, development of objective image quality assessment (IQA) algorithms is necessary to measure the effect of distortion on the visual quality.

Full-reference (FR) IQA algorithms [1] [2] quantify the difference of a distorted image with its reference image. Reduced-reference (RR) methods use only partial information of the reference image in quality assessment. No-reference (NR) IQA methods [3] [4] predict image quality without using any reference image. FR metrics cannot be used when there is no information about the reference image. In such cases, an NR IQA method is needed. However, designing an NR algorithm performs well on different image contents and distortion types is very challenging. RR IQA methods attain a good trade-off between FR and NR metrics as they use partial information of reference image and achieve high performance.

In general, RR IQA models are based on extracting a number of quality-aware features from reference and distorted images. Wang et al. [5] proposed an RR method using a natural image statistic model in the wavelet domain. It computes the Kullback-Leibler distance between marginal probability distribution of wavelet coefficients of reference and test images. Soundararajan et al. [6] proposed an RR IQA that uses the difference of the weighted entropies between the reference and distorted images in the wavelet domain. Based on the orientation selectivity mechanism of the primary visual cortex, Wu et al. [7] proposed an orientation selectivity based visual pattern (OSVP) to extract visual content for RR IQA.

Current IQA methods are mostly devised for evaluation of images degraded by single distortion types. However, different processing steps may introduce multiple distortion types to images. In communication systems, images may pass through acquisition, compression and transmission steps before reaching end users. Therefore, images may be subjected simultaneously to multiple distortion types hence cluttering final IQA. Two new multi-distortion databases namely MLIVE (LIVE multiply distorted image quality database) [8] and MDID2015 (multiply distorted image database) [9] challenge many state-of-art IQA methods. Therefore, it is of great importance to design IQA models for images corrupted by multiple distortion types.

The interaction between various distortion types may complicate the design of an effective metric. IQA design becomes even more challenging in case of RR methods when the reference image is not fully accessible. In this paper, we propose an RR IQA method for quality assessment of images subjected to multiple distortion types. The input images are decomposed into predicted and disorderly parts according to the internal generative mechanism (IGM) [10]. A number of quality-characterizing features are extracted from each part based on the shearlet transform and Rényi directional entropy. Then, the feature difference values are computed between reference and distorted images and the support vector regression (SVR) is adopted to obtain a quality score. The performance of

QoMEX2017 – Erfurt, Germany; 978-1-5386-4024-1/17/$31.00

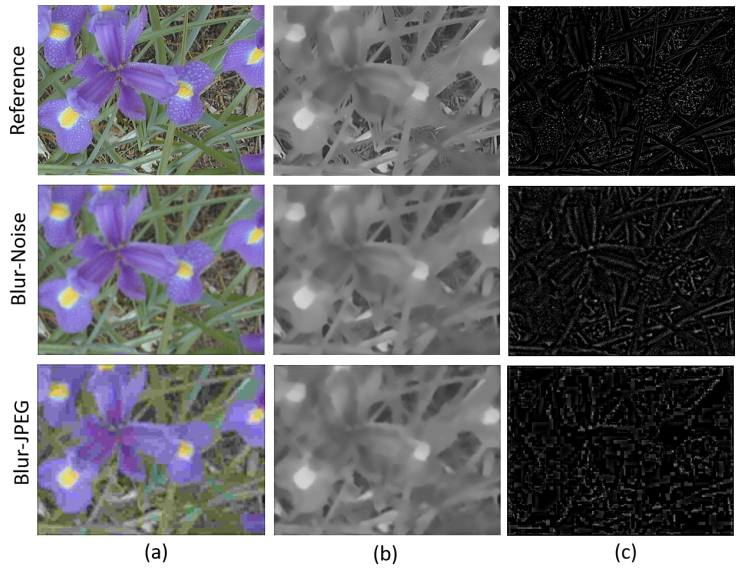

Fig. 1. (a) IGM-based image decomposition: (a) Original (reference) image, (b) predicted portion (I_P) and (c) disorderly portion (I_D) [scaled to 0-255].

the proposed RR-IQA approach is evaluated on the MDID2015 and MLIVE multi-distortion databases.

The remainder of the paper is organized as follows. In Section II, we describe the image decomposition based on the internal generative mechanism (IGM). Section III elaborates on the feature extraction based on shearlets and Rényi entropy analysis for respectively the predicted and disorderly parts. The extraction of the final quality score computed from these elementary measures is detailed in Section IV. Experimental results are provided in Section V. The final section addresses the conclusions of this paper.

II. IGM-based Image Decomposition

Researches in brain science revealed that the human visual system (HVS) possesses an internal generative mechanism (IGM) [10] [11]. IGM actively predicts the visual content and avoids residual uncertainty/disorders. Inspired by IGM theory, for quality assessment, an image can be decomposed into predicted and disorderly portions.

Wu et al. [12] proposed a Bayesian prediction based autoregressive (AR) model to mimic the visual content prediction of IGM and they decomposed an image into its predicted and disorderly portions. Based on IGM, the prediction of the visual content is highly correlated to similarities among nearby pixels. Thus, the value of a central pixel x_c can be predicted by deploying an AR model as follows:

$$v_c = \sum_{v_i \in X} \rho_i v_i + \epsilon \tag{1}$$

where v_c is the predicted value of x_c and v_i is the value of neighbourhood pixels X. The normalized correlation coefficients are represented by ρ and ϵ is a term characterized as white noise.

Using (1), an input image I is decomposed into the predicted image I_P and the disorderly image I_D = I - I_P. Fig. 1a shows a reference image from the MDID database with its predicted and disorderly parts. The predicted image presents the main visual content. The distortions on this part mostly affect the visual structure of an image which can degrade visual understanding. The disorderly image contains residual uncertainty information. Distortions on this part mainly change the disorder of the image, which causes an uncomfortable perception with limited effect on image understanding.

Various distortion types induce different degradations on the predicted and disorderly parts. For example, additive white Gaussian noise (AWGN) has no significant effect on the visual structure and mainly causes uncomfortable perception. Therefore, AWGN is more likely to appear in the disorderly portion. Oppositely, blur changes the primary visual information (edges) and degrades the visual understanding, impacting mainly the prediction portion. Since different distortion types have a distinct effect on the two decomposed portions, we proposed to use the IGM-based image decomposition for quality assessment of multiply distorted images. Fig. 1b and 1c show two input images subjected to multi-distortion types (blur-noise and blur-jpeg). As shown in Fig. 1b, the blur degradation can be better observed in the predicted part while the noise mainly changes the disorderly image. In Fig. 1c, the blur mainly damaged the structural information in the predicted part, while the JPEG degrades both parts.

Designing IQA models for multiple distortion types is challenging since one should consider the interaction between the distortions the image was subjected to. We suggest that the degradation effect of multiple distortion types can be better interpreted on predicted and disorderly images. Inspired by the IGM based prediction model, we proposed a new RR IQA method that quantifies the degradation effect of multiple distortion types on predicted and disorderly parts.

III. Proposed Method

First, reference and distorted images (I and I') are decomposed into predicted (I_P and I'_P) and disorderly (I_D and I'_D) images (Fig. 2). Then, a number of quality-characterizing

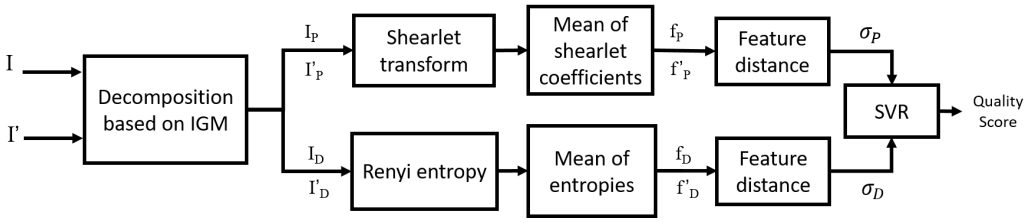

Fig. 2. Overview of the proposed framework

features are extracted from the decomposed images and a quality index is obtained using support vector regression (SVR). The degradation in the predicted part is modelled by feature extraction in the shearlet domain, yielding quality features (f_P and f'_P). The quality features of the disorderly part (f_D and f'_D) are obtained by computing the Rényi directional entropy. Next, the differences between the features of reference and distorted images are separately computed for predicted and disorderly parts. Finally, the difference values are fed in the SVR for quality assessment.

A. Features of the Predicted Image

First, we extract a number of features from the predicted parts of reference and distorted images to obtain their corresponding feature vectors. The predicted part of an image includes primary visual information such as edges. Such information can be well presented in the shearlet domain. The shearlet transform [13] provides a sparse representation for multidimensional data and anisotropic information, most notably edges, at multiple scales. The quality degradation can be predicted by quantifying the deviations of shearlet coefficients of distorted images from those of a reference image. Here, a number of quality-related features are extracted from the predicted images in shearlet domain.

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

$$I \rightarrow SH_{\varphi} I(a,s,t) = \langle I, \varphi_{a,s,t} \rangle \tag{2}$$

where $a>0$ is the scale parameter, $s \in R$ is the shear parameter and $t \in R^2$ denotes the translation parameter. The shearlet coefficient $\varphi_{a,s,t}$ is given by:

$$\begin{aligned} &\varphi_{a,s,t}(x) = |\det M_{a,s}|^{-\frac{1}{2}} \varphi(M_{a,s}^{-1}(x-t)) \text{ where} \\ &M_{a,s} = S_s A_a = \begin{bmatrix} a & s\sqrt{a} \\ 0 & \sqrt{a} \end{bmatrix} \\ &A_a = \begin{bmatrix} a & 0 \\ 0 & \sqrt{a} \end{bmatrix} \qquad S_s = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \end{aligned} \tag{3}$$

$\varphi(.)$ is Meyer wavelet function. To achieve optimal sparsity, the anisotropic dilation matrix A_a ensures the multiscale property, while the shear matrix S_s provides a mean to detect directions.

2) Features Extraction: The predicted part of reference and distorted images are transformed into one low-pass shearlet subband and ten high-pass directional subbands using 1-level shearlet decomposition. Since high frequency components of an image are more sensitive to distortion, ten directional subbands of the finest – i.e. highest frequency – scale are considered for feature extraction. In each directional subband i ($1<i<10$), the mean of shearlet amplitudes is computed as feature. Finally, the normalized difference σ_P between the mean values of reference and distorted images (in i-th subband) is obtained as follows:

$$\sigma_P(i) = \frac{\mu'_P(i) - \mu_P(i)}{\mu_P(i)} \tag{4}$$

where μ'_P and μ_P are the mean values of the shearlet amplitudes in the distorted and reference images, respectively. Using (4), ten quality measures are obtained from the predicted part.

B. Features of the Disorderly Image

The disorderly portion of an image represents information that the HVS cannot interpret. The pixel values in the disorderly image depict the degree of uncertainty. As shown in Fig. 1, each distortion has a distinct impact on the amount and composition of information in the disorderly part. To capture the amount of the information changes in various directions, the generalized Rényi entropy [14] has been employed.

Information entropy is an effective measure of the amount of information in an image and distortion can alter image entropy in different ways. To compute the amount of directional information, Gabarda et al. [14] proposed a generalized Rényi entropy in which the directional selectivity is enabled by using the 1-D pseudo-Wigner distribution (PWD) [15] implementation. The general Rényi entropy is defined as:

$$R_{\gamma} = \frac{1}{1-\gamma} \log_2 \left(\sum_n \sum_k U^{\gamma}[n,k] \right) \tag{5}$$

where $\gamma \geq 0$ and $\gamma \neq 1$, and $\gamma \geq 2$ is recommended for space-frequency distribution measures. $U[n,k]$ represents the discrete space-frequency distribution of the image, and n, k are space and frequency variables, respectively. The spatial-frequency distribution of a given image $U[n,k]$ can be modeled by associating the gray-level spatial data with Wigner spatial-frequency distribution. A discrete approximation of the Wigner distribution [15] is given by:

$$W_z[n,k] = 2 \sum_{r=-L/2}^{L/2-1} z[n+r]\, z^*[n-r] e^{-2i(2\pi r/L)k} \tag{6}$$

where r is a shifting parameter, $z[n]$ indicates a 1-D gray values sequence of L pixels in the desired direction, and z^*

is the complex conjugate of z. The equation is in the spatial interval $[-L/2, L/2-1]$ (the PWD window), which allows local information extraction. Full PWD of the image can be obtained by shifting the PWD window to all possible positions over the image. By rotating the window in different directions, we can obtain directional distributions.

Using a proper normalization, $W_z[n,k]$ is associated to $U[n,k]$ and then a pixel-wise Rényi entropy can be measured.

The PWD is computed in a 1-D oriented window to obtain the entropy in a selected direction. We compute the pixel-wise Rényi entropy in 6 equally-spaced directions ($0\,^\circ$, $30\,^\circ$, $60\,^\circ$, $90\,^\circ$, $120\,^\circ$, and $150\,^\circ$) for both the reference and distorted images. The means of the entropy values in each direction j are obtained as features and the difference σ_D between the mean values of the reference and distorted images is computed:

$$\sigma_D(j) = \frac{\mu_D^{'}(j) - \mu_D(j)}{\mu_D(j)} \tag{7}$$

where μ_D' and μ_D are the mean values of the pixel entropies in the disorderly part of the distorted and reference images, respectively. Hence, six quality measures are obtained from the disorderly part.

IV. Quality Evaluation

Hence, we obtained as such ten quality measures from the predicted part σ_P and six measures from the disorderly part σ_D. In total, sixteen measures are collected for quality assessment. Since the effect of the different distortion types are not similar in the predicted and the disorderly parts, the performance of the sixteen measures in these two parts will differ depending on the natures of the distortions the image was subject to. To combine these quality measures with appriopriate weights, support vector regression (SVR) [16] is adopted and a quality prediction model is constructed. The final quality score Q is computed by:

$$Q = SVR(\sigma_P\ ,\ \sigma_D) \tag{8}$$

In our method, SVR with a radial basis function (RBF) kernel was implemented, utilizing the LIBSVM package [17].

V. Experimental Results

In this section, we examined the ability of the proposed method for quality estimation of multiply-distorted image databases (MDID2015 and MLIVE).

The MDID2015 image database was used for training the IQA model. The database contains 20 reference images corrupted with five distortions of random types and levels. The distortion types are Gaussian noise (GN), Gaussian blur (GB), contrast change (CC), JPEG, and JPEG 2000. The CC distortion is excluded and a total number of 558 distorted images are used in the experiment. The human subjective ratings are reported as differential mean opinion scores (DMOS). A 5-parameter logistic function is used to nonlinearly map the quality scores of objective metrics to the DMOS scores:

TABLE I. Median LCC and SROCC comparison across 100 train-test on MDID2015 database.

	LCC	SROCC
DIIVINE (NR)	0.573	0.498
ShearletIQM (NR)	0.575	0.502
BLIINDS II (NR)	0.584	0.509
SSIM (FR)	0.698	0.676
PSNR (FR)	0.711	0.690
FSIM (FR)	0.899	0.896
HDR-VDP-2.2 (FR)	0.907	0.903
OSVP (RR)	0.743	0.705
Proposed Method (RR)	0.823	0.796

$$f(q) = \alpha_1 \left(\frac{1}{2} - \frac{1}{1 + e^{(\alpha_2 \cdot (q - \alpha_3))}} \right) + \alpha_4 . q + \alpha_5 \tag{9}$$

where *f(q)* is the predicted score and α_1 to α_5 are fitting parameters

The MDID database was iteratively partitioned in train and test data sets for performance evaluation. The train and test images were separated by content to ensure the validity of experiment. The training set contained 80% of the original images and their distorted versions and the remaining 20% of images were included in test set. A regression model is obtained from each training set which is used for quality assessment of test images. The random training-test set split was repeated 100 times, and the median performance indices across 100 experiments were reported for each IQA metric.

The performance indices are the linear correlation coefficient (LCC) and Spearman rank order correlation coefficients (SROCC) between the IQA methods and DMOS scores.

The proposed method is compared with several FR (PSNR, SSIM [1], FSIM [2] and HDR-VDP-2.2 [18]), RR (OSVP-RR [7]) and NR IQA (ShearletIQM [19], DIIVINE [3] and BLIINDS II [4]) metrics.

Table I shows the performance of the proposed method on the MDID2015 database over 100 train-test iterations. Compared to FR methods, the proposed method has higher LCC and SROCC values than PSNR and SSIM. The performance of HDR-VDP is the highest among all methods. The proposed RR method delivers superior performance over the competing RR and NR IQA methods.

The median values in Table I illustrate the performance differences between competing methods. However, these differences may not be statistically relevant. Thus, we further computed the statistical significance of five metrics with highest performance. A two-sample t-test is performed on the SROCC values over 100 train-test trials. The results are shown in Table II. A value '1' in Table II shows that an IQA method on the horizontal axis is statistically superior (with 95% confidence) to a method on the vertical axis. The value is '0' when there is no statistically significant distance between two IQA methods (or the two metrics are equivalent). A value '-1' indicates that the IQA metric on the horizontal axis is statistically inferior to a metric on the vertical axis.

To prove that the proposed method is independent from the test database (MDID2015), we also performed an evaluation

TABLE II. RESULTS OF STATISTICAL SIGNIFICANCE T-TEST ON THE SROCC VALUES.

	PSNR	SSIM	FSIM	HDR-VDP	OSVP	Proposed
PSNR	0	0	-1	-1	-1	-1
SSIM	0	0	-1	-1	-1	-1
FSIM	1	1	0	0	1	1
HDR-VDP	1	1	0	0	1	1
OSVP	1	1	-1	-1	0	-1
Proposed	1	1	-1	-1	1	0

TABLE III. THE LCC AND SROCC COMPARISON ON MLIVE DATABASE.

	LCC	SROCC
BLIINDS II (NR)	0.357	0.248
ShearletIQM (NR)	0.582	0.536
DIIVINE (NR)	0.639	0.608
SSIM (FR)	0.734	0.645
PSNR (FR)	0.737	0.677
FSIM (FR)	0.893	0.863
HDR-VDP-2.2 (FR)	0.896	0.874
OSVP (RR)	0.775	0.732
Proposed Method (RR)	0.806	0.781

test on the MLIVE database. The MLIVE database consists of 450 distorted images and spans two types of distortions: Blur-JPEG and Blur-Noise. The model was trained on the entire MDID database and the trained model was used for quality evaluation of the MLIVE distorted images. The results in Table III confirm that the proposed method has good performance independent of the trained database.

The computational time of the proposed method is compared with several metrics in Table IV. The original matlab code of each algorithm is executed on the first image of MDID database with resolution 512x384 for 10 iterations and the average time is reported. The total computational cost of the method is 10.45 seconds in which the IGM-based decomposition takes 6.64 seconds. The computational cost of the proposed method is reasonable compared to the machine-learning based approaches (such as BLIINDS II and DIIVINE).

VI. CONCLUSION

An RR IQA method is proposed for quality assessment of multiply-distorted images. Inspired by IGM, the reference and distorted images are decomposed into predicted and disorderly parts in which a number of features are extracted. The shearlet transform is used to capture the features in the predicted part. In the disorderly part, directional Rényi entropy values are computed as features. The features differences between reference and distorted images are computed to obtain quality measures. Finally, quality measures are weighted using a support vector machine utilizing a radial basis function kernel. The evaluation results on two multi-distortion databases, namely MDID2015 and MLIVE show that the proposed RR IQA method has high prediction accuracy and its performance is independent from the trained database and it is competitive with state-of-the-art IQA solutions.

TABLE IV. THE AVERAGE COMPUTATIONAL TIME OF DIFFERENT METHODS (IN SEC).

	HDR-VDP	FSIM	OSVP	BLIINDS	DIIVINE	Proposed
Time (sec)	1.49	0.347	0.12	19.35	10.27	10.66

ACKNOWLEDGMENTS

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] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, 2004.

[2] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: a feature similarity index for image quality assessment," *Image Processing, IEEE Transactions on*, vol. 20, no. 8, pp. 2378–2386, 2011.

[3] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *Image Processing, IEEE Transactions on*, vol. 21, no. 8, pp. 3339–3352, 2012.

[4] A. K. Moorthy, and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *Image Processing, IEEE Transactions on*, vol. 20, no. 12, pp. 3350–3364, 2011.

[5] Z. Wang, G. Wu, H. R. Sheikh, E. P. Simoncelli, E. Yang, and A. C. Bovik, "Quality-aware images," *Image Processing, IEEE Transactions on*, vol. 15, no. 6, pp. 1680–1689, 2006.

[6] R. Soundararajan, and A. C. Bovik, "RRED indices: reduced reference entropic dfferencing for image quality assessment," *Image Processing, IEEE Transactions on*, vol. 21, no. 2, pp. 517–526, 2012.

[7] J. Wu, W. Lin, G. Shi, L. Li, and Y. Fang, "Orientation selectivity based visual pattern for reduced-reference image quality assessment," *Journal of Information Sciences*, vol. 351, pp. 18–29, 2016.

[8] D. Jayaraman, A. Mittal, and A. C. Bovik, "Objective quality assessment of multiply distorted images," in *Signals, Systems and Computers, Forty-Sixth Asilomar Conference on*. 2012, pp. 1693–1697.

[9] W. Sun, F. Zhou, and Q. Liao, "MDID: A multiply distorted image database for image quality assessment," *Journal of Pattern Recognition*, vol. 61, pp. 153–168, 2017.

[10] K. Friston, "The free-energy principle: a unified brain theory?," *Journal of Nature Reviews Neuroscience*, vol. 11, no. 2, pp. 127–138, 2010.

[11] K. Friston, and J. Kinler, and L. Harrison, "A free energy principle for the brain," *Journal of Physiology-Paris*, vol. 100, no. 1-3, pp. 70–87, 2006.

[12] J. Wu, W. Lin, G. Shi, and A. Liu, "Perceptual quality metric with internal generative mechanism," *Image Processing, IEEE Transactions on*, vol. 22, no. 1, pp. 43–54, 2013.

[13] W. Q. Lim, "The discrete shearlet transform: a new directional transform and compactly supported shearlet frames," *Image Processing, IEEE Transactions on*, vol. 19, no. 5, pp. 1166–1180, 2010.

[14] S. Gabarda, and G. Cristobal, "Blind Image quality assessment through anisotropy," *Journal of the Optical Society of America*, vol. 24, no. 12, pp. 42–51, 2007.

[15] T. A. Classen, and W. F. Mecklenbrauke, "The Wigner distribution-a tool for time frequency analysis," *Philips Journal of Research*, vol. 35, no. 4, pp. 276–300, 1980.

[16] A. J. Smola, and B. Scholkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 4, no. 3, pp. 199–222, 2004.

[17] C. C. Chang, and C. J. Lin, "LIBSVM: A library for support vector machines," *Intelligent Systems and Technology, ACM Transactions on*, vol. 2, no. 3, pp. 1–27, 2011.

[18] M. Narwaria, R. K. Mantiuk, M. Perreira Da Silva and P. Le Callet, "A calibrated method for objective quality prediction of high dynamic range and standard images," *Journal of Electronic Imaging*, vol. 24, no. 1, pp. 010501-1–010501-3, 2015.

[19] S. Mahmoudpour, and M. Kim, "No-reference image quality assessment in complex-shearlet domain," *Journal of Signal, Image and Video Processing*, vol. 10, no. 8, pp. 1465–1472, 2016.