Opinion-unaware blind image quality assessment based on sparse representation

Maozheng Zhao, Ran Gao, Bo Yang, Aidong men

School of Information and Communication Engineering
Beijing University of Posts and Telecommunications
Beijing, 100876, China
{*maozhengzhao, gaoran, boyang, menad*}*@bupt.edu.cn*

***Abstract*—Opinion-unaware (OU) General purpose blind image quality assessment (BIQA) is an important yet difficult task in computer vision and machine learning. In this paper, we present a sparse representation based quality-unaware BIQA approach which utilizes NSS features from three different domains. At the training stage, we extract quality scores of training image blocks by full-reference image quality assessment method and NSS features of image blocks. The score-feature pairs are learned by the dictionary for sparse representation. At the testing stage, the NSS feature vector of a test image is represented via sparse coding and the final predictive quality score is obtained through weighting quality scores by the sparse coding coefficients. Experiments on the LIVE database show that the proposed algorithm has better performance than the state-of-the-art opinion-unaware method NIQE. And it also has excellent database independence.**

***Index terms*— no-reference image quality assessment (NR IQA), sparse representation, phase congruency**

I. Introduction

Thanks to the easier availability of digital images from the internet and the popularization of wireless digital imaging devices, digital images have been becoming one of the most important way of representing and transmitting information in our everyday life. However, the perceptual quality of digital images may be declined by the process of acquisition, transmission, compression, restoration, etc. For example, images could be distorted by packet loss when transmitted through a congested wireless network. Moreover, in most real life cases the undistorted version of the observed image is not available, thus assessing image quality blindly and automatically is becoming increasingly important [1].

In terms of whether specifying the distortion types of test images, the current blind image quality assessment (BIQA) methods can be classified as distortion specific methods [2], [3], [4] or general purpose methods [5], [6], [7], [8], [9]. The application of distortion specific BIQA methods is therefore limited to test images with known distortion types. Thus the general purpose BIQA methods which do not require prior knowledge of distortion types of test images are more practical in real-world applications. General purpose BIQA methods can be further classified into two categories: opinion-aware (OA) methods [6], [7], [5], [8] and opinion-unaware (OU) methods [10], [9], [11]. OA methods need subjective scores for training while OU methods can build prediction models without human opinions which is obviously a much more challenging task than OA approaches. In this paper we focus on the OU-BIQA methods.

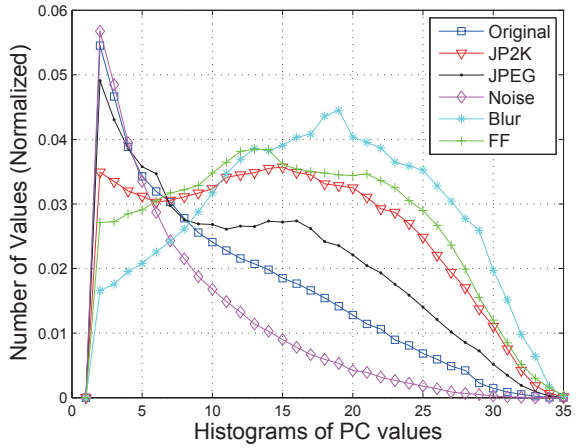
We propose a general-purpose opinion-unaware BIQA approach based on sparse representation (SP). At the training stage of the proposed method, natural scene statistics (NSS) features of the training image blocks and their corresponding quality scores obtained by gradient magnitude similarity deviation (GMSD) [12] which is a state-of-the-art full-reference image quality assessment (FR-IQA) method are computed for dictionary learning. At the testing stage, the NSS feature vector of a test image is represented via sparse coding and the final predictive quality score is computed through weighting GMSD scores by the sparse coding coefficients. Our sparse representation based OU-BIQA approach offers the following advantages. 1) We extract NSS features from three different domains, i.e., phase congruency domain, DCT domain and spacial domain in order to fully employ the quality information of images. 2) The dictionary for sparse representation can be trained completely without human subjective scores since training quality scores are obtained from top performing FR-IQA method. 3) The quality relevant NSS features and the quality scores for constructing the dictionary are extracted from both the whole images and image blocks to make a denser dictionary. 4) By using sparse representation to constuct the dictionary, we do not have to retrain the model when samples of new distortion types are available, we just need to add new samples into the dictionary. 5) At the testing stage, NSS features of the entire image rather than image blocks are used for sparse representation which is highly efficient.

The performance of the proposed opinion-unaware BIQA metric is thoroughly validated on the LIVE IQA database [13]. The experimental results demonstrate that the proposed opinion-unaware BIQA metric is highly consistent with human opinions. It has performance which is very comparable to state-of-the-art OU-BIQA NIQE [9]. It outperforms OU-BIQA QAC [10] and FR-IQA PSNR [14] and its performance is nearly comparable to state-of-the-art OA-BIQA BRISQUE [6]. It also has good extensibility through databases.

The remainder of this paper describes SROUB (Sparse Representation based Opinion-Unaware Blind image quality assessment) in detail and is organized as follows. The details of

This work is supported by the National Science Foundation of China (61271190).

Fig. 2. Histograms of PC for different distortion types. The 6 curves correspond to the reference image "sailing1.bmp" and its distorted counterparts. "Origianl" (DMOS = 0), "JP2K" (DMOS = 67.6968), "JPEG" (DMOS = 70.5024), "Noise" (DMOS = 55.0675), "Blur" (DMOS = 68.0166), "FF" (DMOS = 64.7162).



the proposed framework are described in Section 2. In Section 3, experimental results and a thorough analysis of our results are presented. Finally, Section 4 concludes our work.

II. Sparse representation based opinion-unaware BIQA

Fig. 1 is the framework of SROUB, which includes five key modules, i.e., features and GMSD scores extraction for dictionary, dictionary learning, feature extraction for quality computation, sparse representation and quality computation.

A. Features and GMSD scores extraction for dictionary

In order to get a dense overcomplete dictionary, we extract image features and image GMSD [12] scores not only from each whole image but also from image blocks in order to build a dense dictionary.

We extract image NSS features from three different domains, i.e., phase congruency [15], [16] domain, DCT domain, and spatial domain. The GMSD [12] score of each image block is also computed as the training quality labels of each image block . GMSD is one of state-of-the-art FR-IQA methods whose predictive scores correlate well with human opinions. The Spearman rank order correlation coefficient (SROCC) between GMSD scores and subjective scores on LIVE IQA database [13] is 0.960.

1) PC domain NSS features: Phase congruency (PC) [17], [16] measures the degree of congruency of image local Fourier waves frequencies. The physiological and psychophysical evidences had corroborated that visually cognizable features coincide with points of high PC values [17], [16].

Kovesi [16] proposed an efficient approach of computing PC which is widely used in literature. A 2-D extended logarithmic Gabor filters [18] is adopted for wavelet transform:

$$G_2(\omega,\theta_{\mathrm{k}}) = \exp\left(-\frac{\left(\log(\frac{\omega}{\omega_0})\right)^2}{2\sigma_l^2}\right)\cdot\exp\left(-\frac{(\theta-\theta_{\mathrm{k}})^2}{2\sigma_\theta^2}\right) \tag{1}$$

where $\theta_k = k\pi/K$, $k = \{0,1,...,K-1\}$ is the filter's orientation angle, K is the number of orientations, ω_0 is the center frequency of the filters, σ_l controls the filter's bandwidth and σ_θ determines the filter's angular bandwidth.

By convolving G_2 with an image and modulating ω_0 and θ_k, we get a set of responses at each point i as $\lfloor e_{n,\theta_k}(i), o_{n,\theta_k}(i)\rfloor$, where $e_{n,\theta_k}(i)$ and $o_{n,\theta_k}(i)$ are the real and imaginary parts of a log-Gabor filter respectively on scale n and orientation θ_k at location i. The local amplitude on scale n and orientation θ_k is $A_{n,\theta_k}(i) = \sqrt{e_{n,\theta_k}(i)^2 + o_{n,\theta_k}(i)^2}$, and the local energy along orientation θ_k is $E_{\theta_k}(i) = \sqrt{L_{\theta_k}(i)^2 + T_{\theta_k}(i)^2}$, where $L_{\theta_k}(i) = \sum_n e_{n,\theta_k}(i)$ and $T_{\theta_k}(i) = \sum_n o_{n,\theta_k}(i)$. The 2-D PC at i is as follows:

$$PC(i) = \frac{\sum_k E_{\theta_k}(i)}{\gamma + \sum_n \sum_k A_{n,\theta_k}(i)} \tag{2}$$

where γ is a small positive constant to keep the fraction stable when all the Fourier amplitudes are very small. $PC(i)$ is a real number within $[0,1]$.

We found that the histogram of PC values of one image would change with different distortion types and extents. Fig. 2 shows the histograms of PC values of the image "sailing1.bmp" of the LIVE database [13] and its 5 distortion versions, number of zeros of PC are excluded. We can see that the mean and skew values of PC histograms change with distortion types and extents.

For NSS features from PC domain, the first step is computing pixel based PC values of image blocks. Then by sorting PC values in ascending order, the ordered set $P = (p_1, p_2, ..., p_m)$ is yielded. The percentile pooling is implemented to P based on the assumption that distortions only occur in a subset of pixels. Extracting the central 60% elements from P yields $P_p = (p_{\lfloor 0.2n\rfloor}, p_{\lfloor 0.2n\rfloor+1}, ..., p_{\lfloor 0.8n\rfloor})$. Where n is the number of pixels of the relevant image scale. The mean value of P_p and the skew value of P are the NSS features of PC domain on one scale in our approach.

$$f_{PC} = (mean(P_p), skew(P)) \tag{3}$$

2) DCT domain NSS features: DCT coefficients based on 8×8 nonoverlapping patches is computed first. Then the spectral entropy F is computed based on the DCT coefficients.

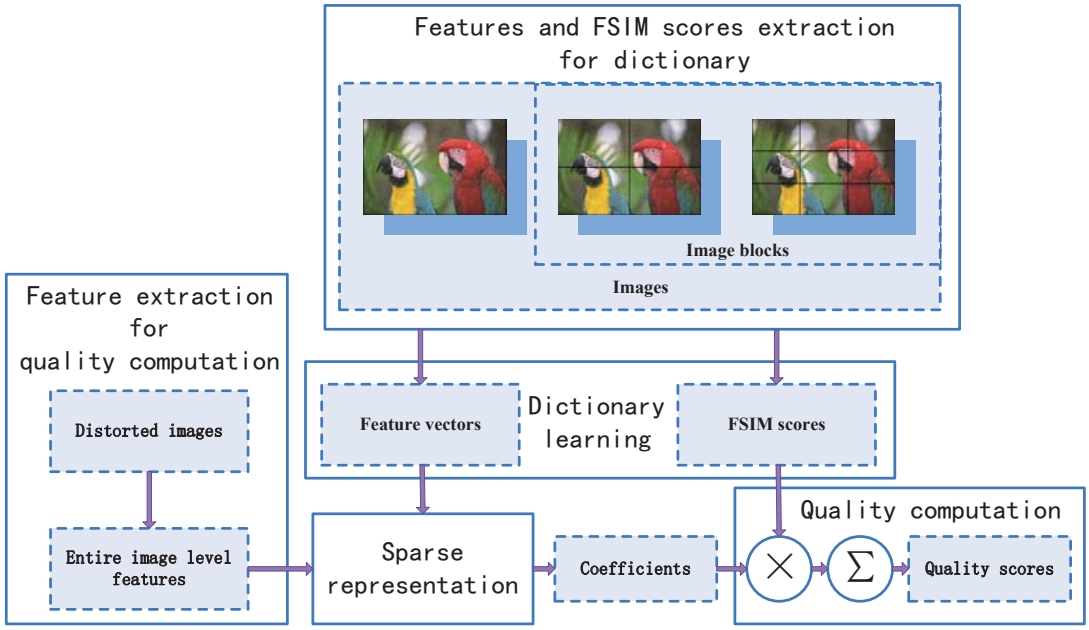
The spectral entropy is

$$F = -\sum_x \sum_y P(x,y)\log_2 P(x,y) \tag{4}$$

where

$$P(x,y) = \frac{C(x,y)^2}{\sum_x \sum_y C(x,y)^2} \tag{5}$$

Fig. 1. Framework of sparse representation based opinion unaware blind image quality assessment.



where $P(x,y)$ is the spectral probability of normalized DCT coefficients $C(x,y)^2$, $1<i\leq 8, 1<j\leq 8$ (DC is excluded).

The same percentile pooling strategy used for PC NSS features is also implemented to the ascending ordered F resulting $Fp=(f_{\lfloor 0.2m\rfloor}, f_{\lfloor 0.2m\rfloor+1}, ..., f_{\lfloor 0.8m\rfloor})$. Where m is the number of blocks within each scale. The mean values of F_p and the skew values of F are used as NSS features of DCT domain in our approach.

$$f_{DCT}=(mean(F_p), skew(F)) \tag{6}$$

3) Spatial domain NSS features: For NSS features from spatial domain, we adopt the same features used in BRISQUE [6] since they require very low time complexity and have strong representing power for image quality. 18 $(2+4\times 4)$ spatial feature vector elements in total are extrated for each scale including parameters of generalized Gaussian distribution (GGD) with zero mean of local mean subtract contrast normalization (MSCN) coefficients and parameters of a zero mode asymmetric generalized Gaussian distribution (AGGD) of products pairs of adjacent coefficients .

B. Dictionary learning

In SROUB, the dictionary learning is extremely simple by directly combining feature vectors and GMSD scores of the training blocks.

$$\begin{bmatrix} D \\ scores \end{bmatrix} \doteq \begin{bmatrix} f_1 & f_2 & ... & f_n \\ score_1 & score_2 & & score_n \end{bmatrix} \tag{7}$$

where the dictionary D is a $k\times l$ matrix ($k=44$ and l is the number of image blocks for training). Vector $scores$ is the corresponding GMSD scores of the training blocks. f_i is the NSS feature vector of the ith training image block and $score_i$ is its GMSD score.

C. Sparse representation

Sparse representation with an overcomplete basis in human perception has been strongly corroborated by studies of the human visual system (HVS) [19]. And the sparsity characteristic is one of the most important features of HVS.

The following optimizing problem can yield the sparsest and the most precise representation of the feature vector of a test image by the atoms in the dictionary $D=[f_1, f_2, ..., f_n]$.

$$\rho^*=\arg\min_{\rho\in R^N} \|\rho\|_0 \quad s.t. \quad f=D\rho \tag{8}$$

where N is the number of atoms in D, ρ^* is the sparse representation coefficients for f over the dictionary D.

The approximate solution of the above optimize problem via l_1-minimization [20] is usually used because the l_0-minimization of equation (8) is an NP hard problem. Then we have

$$\rho^*=\arg\min_{\rho\in R^N} \|\rho\|_1 \quad s.t. \quad f=D\rho \tag{9}$$

This l_1-minimization problem can be transformed into an unconstrained problem

$$\rho^*=\arg\min_{\rho\in R^N} \gamma\|\rho\|_1+\|f-D\rho\|_2 \tag{10}$$

where γ is the regularization parameter. A specialized interior-point method for solving large-scale l_1-regularized least-squares programs is used to solve the above problem [20].

D. Quality computation

Based on the assumption that image blocks with identical quality scores have similar feature distributions, the quality of a test image can be quantified by

$$Q = \frac{\sum_{i=1}^{M} \rho_i^* score_i}{\sum_{i=1}^{M} \rho_i^*} \tag{11}$$

where $score_i$ is the GMSD score of the i-th image block in the dictionary. The estimation Q is the final predictive quality score of the test image.

III. Experiments and results

In the training stage, each image is uniformly partitioned into 4 and 9 neighboring nonoverlapping blocks as shown in Fig. 1. Feature-score pairs are extracted from those blocks which means that each image for building the dictionary would contribute 15 $(1 + 4 + 9)$ feature-score pairs in total. We extract all NSS features at 2 scales to capture multi-scale characteristics via downsampling by a factor of 2 after low pass filtering. For efficiency, only the entire image NSS features are used for sparse representation in the testing stage.

A. Correlation with human opinions

We tested SROUB on LIVE IQA database [13] which consists of 29 reference images and their distorted versions with five different types of distortion, i.e., JPEG2k (JP2K), JPEG, white noise (WN), Gaussian blur (BLUR) and fast fading (FF). DMOS values of those images are offered by the database. We compared performance of SROUB with other IQA approaches. PSNR [14] and SSIM [21] are two classic FR-IQA algorithms. BRISQUE [6] is one of state-of-the-art OA-BIQA methods. QAC [10] and NIQE [9] are state-of-the-art OU-BIQA methods.

Two performance indices were used in our experiments. The Spearman rank order correlation coefficient (SROCC) between DMOS and objective scores is a measure of the prediction monotonicity. The Pearson linear correlation coefficient (LCC) is an evaluation of the prediction accuracy. LCC were calculated after a logistic non-linearity mapping [22].

Since SROUB and BRISQUE require training. We randomly chose 80% of reference images and their corresponding distorted versions in the LIVE database for training and the rest images in the database for testing. This train-test pair is performed 1000 times and the median of the results are reported. For a fair comparison, PSNR, SSIM, QAC and NIQE which do not require training are tested on 20% of randomly chosen reference images and their corresponding distorted versions of the LIVE database for 1000 iterations and the median of the results are reported.

The results are listed in Tab. I. The overall performance across 5 distortion types are shown in the columns labeled by "ALL". We can see that SROUB has higher SROCC and LCC in "ALL" than the FR-IQA PSNR and the two OU-BIQA algorithms QAC and NIQE, its overall performance is slightly inferior to FR-IQA SSIM and OA-BIQA BRISQUE.

B. Variation with dictionary size

In order to investigate the influence of dictionary size to the overall performance of the proposed algorithm, we conduct the $80\%/20\%$ train-test trails for 100 times for different dictionary sizes, the median values of SROCC on the LIVE database are reported. The median values of SROCC are shown in Tab. II, in Tab. II the subscript of a particular dictionary "D" means the number of blocks used for feature extraction, subscript of "1" corresponds to the entire image, subscript of "4" or "9" corresponds to number of adjacent uniformly partitioned blocks of an image used for dictionary construction. From Tab. II, we can conclude that a denser dictionary of the proposed approach results in better correlation with human opinions.

TABLE II
Median SROCC across 100 train-test trials with dictionaries of different sizes on the LIVE IQA database.

	D_1	D_4	D_{1+4}	D_9	D_{1+4+9}
SROCC	0.8361	0.8805	0.8863	0.9058	0.9084

TABLE III
SROCC yielded by models using different features for sparse representation on the LIVE database.

	JP2K	JPEG	Noise	Blur	FF	ALL
Curvelet features	0.7336	0.8891	0.8354	0.8554	0.6792	0.8154
Spatial features	0.4955	0.8681	0.9604	0.8590	0.8100	0.8198
Curvelet & Spatial features	0.7714	0.8683	0.9315	0.8670	0.7553	0.8495

C. Representing powers of features

We also conducted experiments with features from different domains to determine representing powers of features. Here we refer to DCT domain features together with PC domain features as curvelet features. Image blocks in D_1 were used for building dictionaries. We randomly chose 6 reference images (sailing3.bmp, caps.bmp, parrots.bmp, sailing1.bmp, statue.bmp) and their distorted versions from the LIVE database as test images, the rest images in the database are used as training images. The performance of models utilizing only curvelet features or only spatial domain features or both curvelet features and spatial domain features are reported in Tab. III.

The model utilizing both curvelet features and spatial domain features yields relatively high SROCC values on each distortion type and across five distortion types which means that our combination of NSS features from different domains has better representation power for image distortion.

D. Database independence

In order to determine the database independence degree of SROUB, we also test SROUB and the compared algorithms on the TID2008 [23] database. The SROCC values of four distortion types (JPEG2000, JPEG, WN and BLUR) and across four distortion types are listed in Tab. IV. Compared with other approaches, the correlation with subjective perception of SROUB remained consistently competitive.

TABLE I
MEDIAN SROCC AND LCC ACROSS 1000 TRAIN-TEST TRIALS ON THE LIVE IQA DATABASE. ITALICS INDICATE OPINION-UNAWARE BLIND METHODS

	SROCC						LCC					
	JP2K	JPEG	WN	BLUR	FF	**ALL**	JP2K	JPEG	WN	BLUR	FF	**ALL**
PSNR	0.8673	0.8843	0.9428	0.7575	0.8721	0.8633	0.8657	0.8866	0.8963	0.7855	0.8616	0.8568
SSIM	0.9395	0.9480	0.9644	0.9083	0.9413	0.9129	0.9513	0.9605	0.9828	0.9033	0.9569	0.9057
BRISQUE	0.9095	0.9645	0.9778	0.9511	0.8759	0.9363	0.9120	0.9733	0.9848	0.9457	0.9013	0.9385
QAC	0.8849	0.9448	0.9497	0.9181	0.8251	0.8754	0.8691	0.9579	0.9267	0.9228	0.8536	0.8710
NIQE	0.9236	0.9440	0.9720	0.9408	0.8630	0.9089	0.9359	0.9417	0.8712	0.9534	0.8873	0.9082
SROUB	0.9034	0.9592	0.9548	0.9482	0.8912	0.9206	0.9101	0.9635	0.9695	0.9367	0.9331	0.9202

TABLE IV
SROCC OBTAINED BY TRAINING ON THE LIVE IQA DATABASE AND TESTING ON THE TID2008 DATABASE. ITALICS INDICATE OPINION-UNAWARE BLIND METHODS.

	JP2K	JPEG	WN	BlUR	ALL
PSNR	0.7848	0.8585	0.9147	0.8914	0.7527
SSIM(SS)	0.9603	0.9354	0.8168	0.9598	0.9016
BRISQUE	0.9037	0.9102	0.8227	0.8742	0.8977
QAC	0.8885	0.8981	0.7070	0.8504	0.8697
NIQE	0.8940	0.8722	0.7775	0.8236	0.7970
SROUB	0.9221	0.9191	0.8212	0.8987	0.8892

IV. CONCLUSION

We presented an opinion-unaware general purpose BIQA based on sparse representation which utilizes NSS features from three different domains. The quality of a test image is expressed as a simple linear combination of training blocks' GMSD scores with sparse representation coefficients. Via comprehensive experimental validations, we could have the following conclusions. First, SROUB can be trained completely without human subjective scores. Second, SROUB can assess image quality consistently with the subjective scores across various types of distortions. Third, on the LIVE database, SROUB has superior performance to that of the state-of-the-art OU-BIQA method NIQE. Finally, SROUB also has excellent database independence.

REFERENCES

[1] Z. Wang, "Applications of objective image quality assessment methods [applications corner]," *Signal Processing Magazine, IEEE*, vol. 28, no. 6, pp. 137–142, 2011.

[2] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of jpeg compressed images," in *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol. 1. IEEE, 2002, pp. I–477.

[3] J. Caviedes and S. Gurbuz, "No-reference sharpness metric based on local edge kurtosis," in *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol. 3. IEEE, 2002, pp. III–53.

[4] L. Meesters and J.-B. Martens, "A single-ended blockiness measure for jpeg-coded images," *Signal Processing*, vol. 82, no. 3, pp. 369–387, 2002.

[5] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," *Signal Processing Letters, IEEE*, vol. 17, no. 5, pp. 513–516, 2010.

[6] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *Image Processing, IEEE Transactions on*, vol. 21, no. 12, pp. 4695–4708, 2012.

[7] 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.

[8] 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.

[9] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a completely blind image quality analyzer," *Signal Processing Letters, IEEE*, vol. 20, no. 3, pp. 209–212, 2013.

[10] W. Xue, L. Zhang, and X. Mou, "Learning without human scores for blind image quality assessment," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*. IEEE, 2013, pp. 995–1002.

[11] A. Mittal, G. S. Muralidhar, J. Ghosh, and A. C. Bovik, "Blind image quality assessment without human training using latent quality factors," *Signal Processing Letters, IEEE*, vol. 19, no. 2, pp. 75–78, 2012.

[12] X. M. Wufeng Xue, Lei Zhang and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *Image Processing, IEEE Transactions on*, vol. 23, pp. 684 – 695, 2014.

[13] H. Sheikh, Z. Wang, L. Cormack, and A. Bovik, "Live image quality assessment database release 2," 2006, available from: http://live.ece.utexas.edu/research/quality.

[14] Z. Wang and A. C. Bovik, "Mean squared error: love it or leave it? a new look at signal fidelity measures," *Signal Processing Magazine, IEEE*, vol. 26, no. 1, pp. 98–117, 2009.

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

[16] P. Kovesi, "Image features from phase congruency," *Videre: Journal of computer vision research*, vol. 1, no. 3, pp. 1–26, 1999.

[17] M. C. Morrone and D. Burr, "Feature detection in human vision: A phase-dependent energy model," *Proceedings of the Royal Society of London. Series B, biological sciences*, pp. 221–245, 1988.

[18] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *JOSA A*, vol. 4, no. 12, pp. 2379–2394, 1987.

[19] B. A. Olshausen and D. J. Field, "Sparse coding with an overcomplete basis set: A strategy employed by v1?" *Vision research*, vol. 37, no. 23, pp. 3311–3325, 1997.

[20] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale l 1-regularized least squares," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 1, no. 4, pp. 606–617, 2007.

[21] 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.

[22] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *Image Processing, IEEE Transactions on*, vol. 15, no. 11, pp. 3440–3451, 2006.

[23] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "Tid2008-a database for evaluation of full-reference visual quality assessment metrics," *Advances of Modern Radioelectronics*, vol. 10, no. 4, pp. 30–45, 2009.