STRUCTURE-PRESERVING IMAGE QUALITY ASSESSMENT

Yilin Wang[1] *Qiang Zhang*[2] *Baoxin Li*[1]

[1]Department of Computer Science, Arizona State University, Tempe AZ
[2]Advanced Image Research Lab, Samsung Electronic, Pasadena CA
{ywang370,Baoxin.Li}@asu.edu q.zhang1@samsung.com

ABSTRACT

Perceptual Image Quality Assessment (IQA) has many applications. Existing IQA approaches typically work only for one of three scenarios: full-reference, non-reference, or reduced-reference. Techniques that attempt to incorporate image structure information often rely on hand-crafted features, making them difficult to be extended to handle different scenarios. On the other hand, objective metrics like Mean Square Error (MSE), while being easy to compute, are often deemed ineffective for measuring perceptual quality. This paper presents a novel approach to perceptual quality assessment by developing an MSE-like metric, which enjoys the benefit of MSE in terms of inexpensive computation and universal applicability while allowing structural information of an image being taken into consideration. The latter was achieved through introducing structure-preserving kernelization into a MSE-like formulation. We show that the method can lead to competitive FR-IQA results. Further, by developing a feature coding scheme based on this formulation, we extend the model to improve the performance of NR-IQA methods. We report extensive experiments illustrating the results from both our FR-IQA and NR-IQA algorithms with comparison to existing state-of-the-art methods.

***Index Terms*—** Mean Square Error, Image Quality Assessment, kernel method.

1. INTRODUCTION

Perceptual image quality assessment (IQA) has many multimedia applications such as image denoising [1] and image transmission. Based on the degree of reliance on a reference image, IQA models can be divided into three categories: Full Reference IQA (FR-IQA), Reduced Reference IQA (RR-IQA) and Non-Reference IQA (NR-IQA). FR-IQA needs a reference image for estimating the distortion of a target image. Numerous FR-IQA models have been proposed, including those that incorporate image structure information [2], mutual information [3, 4], and wavelet information [5], etc. For RR-IQA, only partial information of a reference image is needed, while NR-IQA models predict image quality without any information from reference images. Recent NR-IQA approaches have employed Bag of Words [6, 7], and DCT transformation [8], etc. In general, existing approaches belong to one of the above three categories and only work for their respective scenario. One objective of this paper is to develop a unifying approach for both FR-IQA and NR-IQA, hence maximizing the applicability of the IQA model.

In parallel with perceptual/subjective IQA, various objective measures have been employed in multimedia. The most widely-used one is Mean Square Error (MSE) or its variants, due to its simplicity and general effectiveness. MSE simply measures the average per-sample difference between two signals. Since MSE is convex and differentiable, it is easy to use optimization approaches for finding solutions to various models based on MSE. Unfortunately, it is well understood (e.g., [9]) that pixel-wise MSE (or its variants) is not a good measure for perceptual quality, primarily due to the fact that no structural information of the image is considered in computing MSE. Some attempts have been tried to remedy this. For example, in [10], a "perceptual-aware" MSE was proposed by adding Gaussian filter or gradient operator, which helped to improve the correlation between the human perceptual score and the objective metric.

In this work, we aim at building a structure-preserving MSE (SPMSE) which not only retains the computational efficiency and nice mathematical properties of MSE but also leads to the development of effective IQA metrics. Our new formulation of MSE is developed by employing the 'kernel trick': we use HOG feature [11], an effective and efficient descriptor for describing object structures, as an quality kernel between two images, and show that the resultant formulation leads to many of the desired properties. For FR-IQA, experimental results show that SPMSE performs statistically better than the well-known SSIM method and leads to very competitive performance compared with with other state-of-the-art methods on three benchmark datasets.

Moreover, we show that the proposed SPMSE can be employed in recent Bag-of-Words based models for NR-IQA [7, 6]. In such existing models, the distortion image is represented by a feature vector which is the coefficients under a codebook. However, most existing approaches focus on designing hand-crafted features to be used by the codebook training, rather than optimal representation under the codebook. In other words, the feature encoding step, which should contribute to the final quality metric significantly, has been largely ignored. In this work, we propose and empirically compare several coding schemes for NR-IQA based on the SPMSE framework, and show that, compared to vector quantization or sparse coding, the proposed method, structure preserving coding, is more effective for NR-IQA models.

The rest of the paper is organized as follows: Section 2 reviews the related work. Section 3 describes SPMSE for FR-IQA in details and introduces our SPMSE encoding for NR-IQA. In Section 4, experimental results on widely-used datasets are reported; and finally in Section 5, conclusions are made and future improvements and issues are discussed.

2. RELATED WORK

We review the related work on FR-IQA and recent advanced methods in NR-IQA.

FR-IQA: One of the most widely-used and influential FR-IQA method is Structure SIMilarity Index (SSIM). It is based on the assumption that the underlying image quality score is highly related to the image structure. For a pair of a reference image s and a distortion image t, SSIM compares them with image luminance, contrast and structure as: $SSIM(s,t) = \frac{(2\mu_s\mu_t+C_1)(2\sigma_{st}+C2)}{(\mu_s^2+\mu_t^2+C1)(\sigma_s^2+\sigma_t^2+C2)}$, where, for image i,j, μ_i is the local mean intensity, σ_i is the local variance and σ_{ij} is the local covariance. Visual Information Fidelity (VIF) is another IQA approach that captures the signal statistics for image fidelity assessment. In [12], it was argued that the HSV space is appropriate for full-reference image quality assessment, owing to distinctive features of high-quality and low-quality images in this space. In [13], the author provides a gradient similarity method for image quality assessment. For a thorough survey of modern IQA development, please refer to [14]. In contrast to the methods discussed above, the proposed framework starts from the widely used MSE and applies kernel method to the objective function for preserving image structure.

NR-IQA: When images are transferred to some specific domain, e.g., the DCT domain, local descriptors may be modeled by some parametric distribution, based on this, some previous works [8, 15, 16] on NR-IQA have focused primarily on Natural Scene Statistics (NSS). On the other hand, inspired by the success of Bag-of-Words approaches in computer vision, [7] uses visual codebook to assess image quality. Quality measure of a new image is obtained by computing the average of quality scores of the codewords, weighted by their distances to visual words in the image. However, the method requires a large number of codewords and precomputed Gabor-filters. Other than hand-crafted features, [6] proposes an unsupervised feature-learning method based on raw image patches. The proposed method is similar to [7] in term of its codebook-based encoding. However, our goal is to learn the features based on raw image patches for both non-distortion images and distortion images.

3. THE PROPOSED SPMSE FRAMEWORK

3.1. Structure Persevering Mean Square Error

Given two signals $s, t \in R^N$, the objective function of MSE is $||s-t||_2^2/N$. In SPMSE, we introduce a non-linear structure extractor term for each signal as $\frac{1}{N}||\phi(s)-\phi(t)||_2^2$ where ϕ is a mapping function, which maps the original data space to a new feature space. The objective function of SPMSE is:

$$\begin{aligned} SPMSE(s,t) &= \frac{1}{N}||\phi(s)-\phi(t)||_2^2 \\ &= \frac{1}{N}(\langle\phi(s)\phi(s)\rangle - 2\langle\phi(s)\phi(t)\rangle + \langle\phi(t)\phi(t)\rangle) \end{aligned} \tag{1}$$

In Eq. 1, the SPMSE is guaranteed to be non-negative, and thus it can be viewed as a distance measure. Introducing a kernel operation, we can re-write SPMSE as:

$$SPMSE(s,t) = \frac{1}{N}(K(s,s) - 2\times K(s,t) + K(t,t)) \tag{2}$$

where K is a valid Mercer kernel [18], which can be viewed as a non-linear feature similarity measure for the signals. In the next section we will discuss how to choose K.

3.1.1. Kernel Selection

As definition in [18]: "a kernel is a function that returns the inner product between the images of two inputs in some feature space". The intuition of the kernel method is to measure the similarity between two data vectors in a new feature space. The most widely used kernels for images (signals) are polynomial kernels and RBF kernels [18]. A polynomial kernel is given by $K(x,y) = (\langle x,y\rangle + R)^d$ where R and d are kernel parameters. The RBF kernel is defined as $K(x,z) = exp^{\frac{||x-y||^2}{-2\sigma^2}}$. However, trivially bringing them to the proposed objective function is not a good choice for FR-IQA, since the resultant MSE-based kernel function is still based on pixel-wise computation and hence losing the sight of the image structure distortion, which has been argued to be an essential factor for IQA [4, 2].

Thus, one of our goals is to find a proper kernel that helps to retain structural information of an image. Inspired by its success in object detection, e.g. [11], we employ Histogram of Oriented Gradient(HOG) as image quality descriptor. HOG is one of the most-used low-level vision features for object detection and recognition, and the essential thought behinds HOG is that the local appearance and structure in images can be described by its gradient distribution. Based on the following theorem, we show it can be incorporated into our proposed framework as a valid kernel function.

Theorem 1: *HOG operator is a valid kernel function.*

Proof: Let θ_i and M_i be the orientation and magnitude of gradient at pixel i. Then the HOG feature of each pixel is represented by a hard binning indicator.

$$\delta_{i_n} = \begin{cases} 1 & \text{if } \frac{\theta_i}{2\pi} = n - 1 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

For each image block P, the oriented gradient is represented as $\sigma(P) = \sum_{i \in P} M_i \cdot \delta_i$. When measuring the similarity between patches from two different images, it is equivalent to match the patches in the feature space. Thus, we can represent the similarity between image patches in the feature space with a linear kernel:

$$\begin{aligned} K(P, Q) &= \sum_{i \in P} M_i \cdot \delta_i \sum_{i' \in Q} M_{i'} \cdot \delta_{i'} = \sum_{i \in P} \sum_{i' \in Q} M_i \delta_i^T \delta_{i'} M_{i'} \\ &= \sum_{i \in P} \sum_{i' \in Q} M_i M_{i'} \delta_i^T \delta_{i'} \end{aligned} \tag{4}$$

where P, Q are two patches from two images. Since $M_i M_{i'}$ is a non negative scalar and $\delta_i^t \delta_{i'}$ is the inner product of two vectors, then we can substitute two linear kernel $K_M(i, i') = M_i \cdot M_{i'}$, $K_\delta(i, i') = \delta_i^T \delta_{i'}$ in Eq. 4. Thus, $K(P, Q)$ is a valid kernel [18] and it provides a kernel view of HOG.

It is worth noting that, in contrast to [13], where the metric is simply based on the similarity of gradient value from two signals. In the proposed framework, inspired by the success of using HOG for object detection, we utilize the property of the HOG for image structure description. Specifically, in **Theorem 1**, $K_M(i, i')$ measures the similarity of gradient magnitude of two pixels and $K_\delta'(i, i')$ measures the similarity of gradient orientations of two pixels. Thus, instead of measuring the pixel similarity in MSE, the proposed SPMSE can be viewed as a structure similarity measure for image patches (e.g. 8×8 rectangles in HOG).

3.2. Structure-Persevering Coding

In Section 3.1, we proposed a SPMSE framework for the FR-IQA problem, which captures image local structures instead of measuring the pixel-wise error. Noting Eq. 1 is convex and differentiable, we can easily build an objective function to minimize. We now show how the idea may be extended to handle NR-IQA problems.

In recent NR-IQA approaches [7, 6, 15, 17], different features have been designed. However, feature coding has been largely ignored. In other words, how to efficiently encode the features for NR-IQA is still not well addressed. In [7], hard vector coding was used, and in [6], the authors argue soft coding is better, while [17] argues sparse coding is more efficient. In [19], locality linear coding is proposed, the authors observed that the non-zero coefficients are often assigned to the nearby bases of the encoded data. Based on these observations, we compare different coding schemes and then propose a novel feature coding scheme for NR-IQA, which supports feature learning with the proposed SPMSE metric.

Let X be a set of M-dimensional feature vectors extracted from images, $i.e., X = [x_1, x_2, ..., x_N] \in R^{M \times N}$. $C = [c_1, c_2, ..., c_N]$ is a set of code coefficients for X based on codebook $B = [b_1, b_2, ...b_K] \in R^{M \times K}$. C can be generated by different coding schemes for image representation. Based on [19, 17, 20], the proposed coding scheme solves the following optimization problem:

$$\underset{c}{\operatorname{argmin}} \quad \sum_{i=1}^{N} ||x_i - Bc_i||_2^2 + \lambda ||D_i c_i||_2^2 + \mu |c_i| \tag{5}$$

where $D_i \in R^{K \times K}$ is a diagonal matrix, with each element in the diagonal representing the SPMSE score of input image patch i and code basis j, $i.e. D_{i(j,j)} = SPMSE(x_i, b_j)$. The second term in Eq. 5 gives the input patch freedom to decide proportion of similar structure bases in the codebook, while the third term is the sparse regularization term which makes the nonlinear representation of the features. Unfortunately, the above objective function is computational expensive. To alleviate this, we propose an approximation scheme by relaxing the sparse term in the objective function to $\mathbf{1}^T c_i = 1, \forall i$, which still achieves sparsity if we set small values in the solution to zero. Since the Eq. 5 can be decomposed, the encoded feature c_i can be obtained by solving the following optimization problem:

$$\begin{aligned} &\underset{c}{\operatorname{argmin}} \quad \mathcal{J} = ||x_i - Bc_i||_2^2 + \lambda ||D_i c_i||_2^2 \\ &\text{subject to} \quad \mathbf{1}^T c_i = 1; \end{aligned} \tag{6}$$

The Lagrangian function of Eq. 6 is:

$$\mathcal{L} = ||x_i - Bc_i||_2^2 + \lambda ||D_i c_i||_2^2 + \tau(\mathbf{1}^T c_i - 1) \tag{7}$$

where τ is Lagrangian multiplier. Taking the derivation of L and setting it to zero, we can obtain:

$$\tau \mathbf{1} = 2B^T x_i - 2B^T B c_i - 2\lambda D_i^2 c_i \tag{8}$$

Trick 1: $\mathbf{1}^T c_i = 1$ and $x_i^T B c_i$ is a scalar, Eq. 8 can be written as:

$$\begin{aligned} \tau \mathbf{1} + 2x_i^T B c_i \mathbf{1} = 2B^T x_i \mathbf{1}^T c_i - 2B^T B c_i - 2\lambda D_i^2 c_i \\ + 2\mathbf{1} x_i^T B c_i \end{aligned} \tag{9}$$

$$\Rightarrow -\frac{1}{2}(\tau + x_i^T B c_i)\mathbf{1} = (B^T B - B^T x_i \mathbf{1}^T + \lambda D_i^2 - \mathbf{1} x_i^T B) c_i \tag{10}$$

Trick 2: $x_i^T x_i \mathbf{1}^T c_i$ is a scalar, thus it can be added on the both sides:

$$\begin{aligned} -\frac{1}{2}(\tau + x_i^T B c_i)\mathbf{1} + x_i^T x_i \mathbf{1}^T c_i \mathbf{1} = (B^T B - B^T x_i \mathbf{1}^T \\ + \lambda D_i^2 - \mathbf{1} x_i^T B) c_i \\ + \mathbf{1} x_i^T x_i \mathbf{1}^T c_i \end{aligned} \tag{11}$$

Table 1. PLCC comparison of different FR-IQA models

	LIVE(779 images)	TID2008(1300 images)	CSIQ(750 images)	
MSE	0.8739	0.7649	0.8882	0.8279
SSIM	0.9451	0.8530	0.9188	0.8962
VIF	0.9604	0.8938	0.9321	**0.9226**
IFC	0.9268	0.8007	0.8912	0.8599
MAD	0.9394	0.8306	0.8881	0.8762
SPMSE	0.9364	0.8876	0.9213	**0.9096**

$$\Rightarrow -\frac{1}{2}(\tau + x_i^T B c_i - 2x_i^T x_i \mathbf{1}^T c_i)\mathbf{1} = (B^T B - B^T x_i \mathbf{1}^T + \lambda D_i^2 - \mathbf{1} x_i^T B + \mathbf{1} x_i^T x_i \mathbf{1}^T) c_i \tag{12}$$

Finally, the closed form solution can be obtained after normalization as :

$$\begin{aligned} \widetilde{c_i} &= ((B^T - \mathbf{1}x_i^T)(B^T - \mathbf{1}x_i^T)^T + \lambda D_i^2) \setminus \mathbf{1} \\ c_i &= \widetilde{c_i} / \mathbf{1}\widetilde{c_i} \end{aligned} \tag{13}$$

Compared to hard vector quantization encoding [7] which represents images from a single basis, the approximation scheme of Eq. 5 will achieve much smaller error because of the use of multiple bases (soft coding). It is worth noting that the method in [19] is based on pixel-wise representation, lacking the structural information captured by our SPMSE-based scheme. Moreover, we empirically observed that the coding results from [17] tend to select codebook bases that were from images under different distortions, while code bases from our approach tend to belong to images of similar distortion.

4. EXPERIMENTS

4.1. FR-IQA Evaluation Protocol

Database for FR-IQA evaluation: To evaluate the proposed framework, we tested it on three benchmark IQA datasets: LIVE[21], TID2008[22], and CSIQ[12]. The images in these datasets are generated with different type of distortion and associated with human/subjective opinion score. The LIVE database contains 29 reference images and 779 distorted images with 5 different distortions: JPEG2000 compression (JP2K), JPEG compression (JPEG), additive white noise(AWN), Gaussian blur (GB), and Fast fading (FF). The TID2008 database contains 25 reference images and 1700 distorted images with 17 different noise types. Since the last four distortions (totally 400 images) are not structure distortions, e.g. intensity shift, which is a highly subjective task for people to distinguish with, we reported the results on first 13 distortions. This protocol also has been used in [6, 17]. The CSIQ contains 30 reference images and 866 distorted images generated from JP2K, JPEG, AWN, GB, and pink Gaussian noise, the contrast change is also not the structure distortion for us to deal with. Thus, the number of images from CSIQ is 750.

Evaluation: We evaluate the performance of different methods using Pearson Linear Correlation Coefficient (PLCC) and Spearman Rank Order Correlation Coefficient (SROCC). PLCC is considered as a measurement of the prediction accuracy and SROCC is viewed as an evaluation of how well the relationship between the predicted score and the subject opinion score can be described. A good IQA model should have high PLCC and SROCC.

4.2. NR-IQA Evaluation Protocol

Database for NR-IQA evaluation and codebook construction: We use LIVE database for evaluation and adopt CSIQ databse for codebook construction based on the following reasons: First, there is no overlap between CSIQ dataset and LIVE dataset. Second, both CSIQ and LIVE contain four types of distortion: JP2K, JPEG, GB, AWN. Thus, it is reasonable to use codebook generated from CSIQ to represent the images in LIVE instead of TID2008 which has much more noise types than CSIQ. For each image in CSIQ, we randomly extract 10000 7 by 7 raw patches, then using K-means clustering to generate the codebook. In our experiment, the codebook is fixed by 10000×49. This protocol is also used in [7].

NR-IQA Regression and Evaluation: The predicted score is calculated from linear support vector regression (SVR) directly. Since codebook is constructed from unlabeled data, in LIVE database, we randomly pick 80% images associate with human subject score to train the SVR and remaining 20% for testing. Moreover, we repeat the train-test scheme 100 times for cross-validation. It is worth noting that both the training set and the testing set only contain the distorted images. Finally, we use max-pooling to represent image feature.

4.3. Comparison with FR-IQA and NR-IQA algorithms

In this sub-section, we first compare the results of the proposed method with state-of-the-art FR-IQA models including SSIM [2], VIF[3], IFC[4], MAD [12]. Table 1 and Table 2 list the results of SROCC and PLCC of different FR-IQA models respectively. The results are reported from the original papers with default parameter settings. It is worth noting that PLCC results are reported after logistic regression (Eq. (14) and Eq. 15) between predicted score and subject opinion score, which follows the instruction reported in [23].

From Table 2 and Table 1, we can draw the following conclusions. First, the proposed method outperforms a large margin to MSE and is superior to SSIM. Second, the proposed method is comparable to other state-of-the-art method, e.g., VIF, MAD, in terms of average resuls among three benchmark datasets. Moreover, in Table 3, we compare the speed[1] of the proposed method and other top 3 FR-IQA metrics. It can be seen that the proposed is efficient in terms of computation time.

Table 3. Speed Comparison with Top 3 metrics in FR-IQA to MSE

	MSE	SSIM	SPMSE	VIF	MAD
Time(s)	0.0021	0.031	0.043	0.974	2.07
ratio to MSE	1	15	20	458	986

In Table 4 and Table 5, we report the results of our encoding scheme with comparison of state-of-the-art NR-IQA methods and other encoding schemes. The compared methods including BIQI [7], CORINA[7], DIIVINE [16] and BLIINDS (SVM) [8] and we also compared our encoding methods with hard encoding (HC), sparse encoding (SC) [20] and locality linear encoding (LLC) [19]. From the result we can see that our proposed achieves best result among all the encoding schemes which have same codebook, meanwhile, our result is comparable to state-of-the-arts models, e.g., CORINA. Noting that the evaluation of the proposed method only employs general procedures of BOW, the results can be further improved by employing a more powerful regressor (e.g. random forest) or using precomputed features (e.g., NSS) instead of raw image patches.

$$Quality(x) = \beta_1 logistic(\beta_2, (x - \beta_3)) + \beta_4 x + \beta_5 \quad (14)$$

$$logistic(\tau, x) = \frac{1}{2} - \frac{1}{1 + exp(\tau x)} \quad (15)$$

Table 4. SROCC comparison of different NR-IQA models on LIVE

Method	JP2K	JPEG	AWN	GB	FF	ALL
PSNR	0.872	0.885	0.941	0.764	0.875	0.867
SSIM	0.939	0.946	0.965	0.909	0.941	0.913
BIQI	0.856	0.786	0.972	0.910	0.762	0.819
CORINA	0.943	0.955	0.976	0.969	0.906	**0.942**
DIVIINE	0.913	0.910	0.984	0.921	0.863	0.916
BLIINDS	0.929	0.955	0.956	0.923	0.889	**0.931**
SPMSE	0.936	0.948	0.952	0.958	0.872	**0.930**
LLC	0.921	0.941	0.942	0.932	0.862	0.909
HC	0.919	0.948	0.945	0.908	0.905	0.917
SC	0.926	0.958	0.952	0.941	0.852	0.921

Table 5. PLCC comparison of different NR-IQA models on LIVE

Method	JP2K	JPEG	AWN	GB	FF	ALL
PSNR	0.873	0.874	0.928	0.774	0.869	0.855
SSIM	0.920	0.955	0.982	0.891	0.939	0.906
BIQI	0.809	0.901	0.954	0.829	0.733	0.821
CORINA	0.951	0.965	0.987	0.968	0.917	**0.935**
DIVIINE	0.922	0.921	0.988	0.923	0.888	0.917
BLIINDS	0.935	0.968	0.980	0.938	0.896	**0.930**
SPMSE	0.947	0.951	0.971	0.970	0.899	**0.934**
LLC	0.931	0.941	0.943	0.942	0.872	0.919
HC	0.921	0.950	0.965	0.929	0.883	0.917
SC	0.929	0.965	0.959	0.945	0.892	0.925

5. DISCUSSION AND FUTURE WORK

We proposed a simple yet effective approach for image quality assessment. First, we proposed a structure-preserving MSE-like error function for FR-IQA, and the experimental results show that our method is competitive with respect to the state-of-the-art methods and in particular, outperforms the well-known SSIM. Second, we showed that the proposed approach can be applied to the NR-IQA framework as well, through incorporating it in a coding scheme. Even with only a fixed and unoptimized codebook, the experimental results still showed performance comparable to current approaches. Future efforts include at least two possible exentions: a learning-based method for selecting a kernel function more efficiently, and codebook learning for improved NR-IQA.

[1] All the codes are implemented by Matlab and obtained from original authors' webpage. The HOG computation part is written in C and compiled by Matlab.

6. ACKNOWLEDGMENT

Yilin Wang and Baoxin Li were supported in part by a grant (#1135616) from the National Science Foundation (NSF). Any opinions expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

Table 2. SROCC comparison of different FR-IQA models

	LIVE(779 images)	TID2008(1300 images)	CSIQ(750 images)	Weighted
MSE	0.8756	0.7118	0.9060	0.8362
SSIM	0.9479	0.8742	0.9247	0.8946
VIF	0.9636	0.8731	0.9282	**0.9130**
IFC	0.9259	0.7589	0.8827	0.8383
MAD	0.9438	0.8694	0.9604	**0.9142**
SPMSE	0.9564	0.8887	0.9353	**0.9179**

7. REFERENCES

[1] Keigo Hirakawa and Thomas W Parks, "Image denoising using total least squares," *Image Processing, IEEE Transactions on*, vol. 15, no. 9, pp. 2730–2742, 2006.

[2] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, 2004.

[3] Hamid R Sheikh and Alan C Bovik, "Image information and visual quality," *Image Processing, IEEE Transactions on*, vol. 15, no. 2, pp. 430–444, 2006.

[4] Hamid R Sheikh, Alan C Bovik, and Gustavo De Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *Image Processing, IEEE Transactions on*, vol. 14, no. 12, pp. 2117–2128, 2005.

[5] Damon M Chandler and Sheila S Hemami, "Vsnr: A wavelet-based visual signal-to-noise ratio for natural images," *Image Processing, IEEE Transactions on*, vol. 16, no. 9, pp. 2284–2298, 2007.

[6] Peng Ye, Jayant Kumar, Le Kang, and David Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 1098–1105.

[7] Peng Ye and David Doermann, "No-reference image quality assessment based on visual codebook," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*. IEEE, 2011, pp. 3089–3092.

[8] Michele A Saad, Alan C Bovik, and Christophe 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] Zhou Wang and Alan 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.

[10] Wufeng Xue, Xuanqin Mou, Lei Zhang, and Xiangchu Feng, "Perceptual fidelity aware mean squared error," .

[11] Navneet Dalal and Bill Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. IEEE, 2005, vol. 1, pp. 886–893.

[12] Eric C Larson and Damon M Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, pp. 011006–011006, 2010.

[13] Anmin Liu, Weisi Lin, and Manish Narwaria, "Image quality assessment based on gradient similarity," *Image Processing, IEEE Transactions on*, vol. 21, no. 4, pp. 1500–1512, 2012.

[14] Weisi Lin and C-C Jay Kuo, "Perceptual visual quality metrics: A survey," *Journal of Visual Communication and Image Representation*, vol. 22, no. 4, pp. 297–312, 2011.

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

[16] Anush Krishna Moorthy and Alan Conrad 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.

[17] Lihuo He, Dacheng Tao, Xuelong Li, and Xinbo Gao, "Sparse representation for blind image quality assessment," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 1146–1153.

[18] John Shawe-Taylor and Nello Cristianini, *Kernel methods for pattern analysis*, Cambridge university press, 2004.

[19] Jinjun Wang, Jianchao Yang, Kai Yu, Fengjun Lv, Thomas Huang, and Yihong Gong, "Locality-constrained linear coding for image classification," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 3360–3367.

[20] Honglak Lee, Alexis Battle, Rajat Raina, and Andrew Y Ng, "Efficient sparse coding algorithms," .

[21] Hamid R Sheikh, Muhammad F Sabir, and Alan 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.

[22] Nikolay Ponomarenko, Vladimir Lukin, Alexander Zelensky, Karen 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.

[23] Video Quality Experts Group et al., "Final report from the video quality experts group on the validation of objective models of video quality assessment," *VQEG, Mar*, 2000.