LEARNING QUALITY-AWARE FILTERS FOR NO-REFERENCE IMAGE QUALITY ASSESSMENT

Zhongyi Gu, Lin Zhang, Xiaoxu Liu, Hongyu Li*

School of Software Engineering
Tongji University
Shanghai 201804, China

Jianwei Lu

Institute of Advanced Translational Medicine
& School of Software Engineering
Tongji University, Shanghai 201804, China

ABSTRACT

With the rapid development of the usage of digital imaging and communication technologies, there appears to be a great demand for fast and practical approaches for image quality assessment (IQA) algorithms that can match human judgements. In this paper, we propose a novel general-purpose no-reference IQA (NR-IQA) framework by means of learning quality-aware filters (QAF). Using these filters for image encoding, we can obtain effective image representations for quality estimation. Additionally, random forest is used to learn the mapping from feature space to human subjective scores. Extensive experiments conducted on LIVE and CSIQ databases demonstrate that the proposed NR-IQA metric QAF can achieve better prediction performance than all the other state-of-the-art NR-IQA approaches in terms of both prediction accuracy and generalization capabilities.

Index Terms— NR-IQA, natural scene statistics, sparse filtering, random forest

1. INTRODUCTION

Objective image quality assessment (IQA) refers to automatic quality assessment of an image consistent with human perception. With the increasing usage of digital imaging, IQA becomes an essential yet challenging problem. Depending on the availability of non-distorted reference image, IQA approaches can be classified into three categories, full-reference IQA (FR-IQA), reduced-reference IQA (RR-IQA) and no-reference IQA (NR-IQA) [1]. In this paper, we only focus on addressing the NR-IQA problem.

1.1. Related work

Most early NR-IQA algorithms assume that the distortion type is known as prior knowledge, which makes the NR-IQA problem easier and limits the application scope of the approaches. These methods mainly measure the impact of one distortion type on image quality such as blocking [2], ringing [3], blur [4] and compression [5, 6]. In contrast, the goal of general purpose non-distortion-specific (NDS) NR-IQA approaches is to predict the quality without prior knowledge of the distortion type.

Most existing NDS NR-IQA approaches can be classified into two categories, opinion-aware and opinion-unaware approaches.

Opinion-aware approaches need a collection of distorted images associated with their subjective scores to train a model, which then can be used to estimate the quality of new-coming distorted images. In [7], Moorthy *et al.* proposed a two-step framework, namely BIQI, which first classified one image into one distortion category and then used distortion-specific quality metric to predict the quality. Later, Moorthy *et al.* proposed another NR-IQA metric, DIIVINE [8], which was an extension of BIQI. The deficiency of these two methods is that they assume the distortion type is already contained in the training images. Except for the two-step framework, opinion-aware approaches mainly follow two trends [9], natural scene statistics based (NSS-based) and training-based methods. The design rationale of NSS-based approaches is that the existence of distortion on the image will affect certain statistical properties of natural scenes. In [10], Sadd *et al.* proposed an NR-IQA model, namely BLIINDS, by assuming that the statistics of DCT features would vary in a predictable way as the image quality changes. Later, they improved BLIINDS to obtain an extension, namely BLIINDS-II [11]. In [12], Mittal *et al.* used locally normalized luminance coefficients in the spatial domain to predict the image quality based on the observation that presence of distortion would affect the regular structure of this image coefficients. The goal of training-based methods is to design quality-relevant features that can capture the factors on which the distortion may have impact on. Most of these training-based approaches need to design a large number of hand-craft features. For examples, In [9], Ye and Doermann proposed a codebook-based framework, which is commonly applied to image classification, to learn the regression model. This method was referred to as CBIQ. Later, Ye *et al.* improved CBIQ by using features learned by unsupervised feature learning to re-

*Corresponding author. Email: cslinzhang@tongji.edu.cn

place hand-craft features extracted by Gabor filter. This NR-IQA metric, namely CORNIA [13], was proved to be effective dealing with lots of distortion types.

Opinion-unaware approaches have the advantage that they do not require training on databases associated with human scores. For examples, in [14], Mittal *et al.* proposed a method by conducting probabilistic latent semantic analysis on the statistical features of a large collection of pristine and distorted image patches. In [15], Xue *et al.* at first used a quality-aware clustering method to learn centroids from images of different quality levels, and then use these centroids to infer image quality. One innovation of their method is that the image scores for clustering were calculated by a FR IQA metric, FSIM [16]. To the best of our knowledge, the model proposed in [17], namely NIQE, is the most accurate opinion-unaware approach in current literature. NIQE learned a MVG model from pristine images and estimated the quality of a distorted image by measuring the distance between multivariate Gaussian (MVG) fit of the distorted image and the pristine images. According to the experiments conducted on LIVE IQA database [18], the prediction accuracy of opinion-unaware methods is lower than that of opinion-aware ones.

1.2. Our approach

As an alternative, in this paper, we propose a general-purpose opinion-aware NR-IQA method by learning quality-aware filters (QAF). The keys to QAF are sparse filtering [19] and random forest [20]. Sparse filtering is typically used to learn a filter dictionary, which maps the original data to good feature representations for classification tasks by optimizing exclusively for sparsity in the feature distribution. Here we apply sparse filtering to a set of NSS-based features extracted from image patches of different quality degrees to learn a quality-aware filter dictionary. The term "quality-aware" means that one particular filter in the dictionary only gives strong response to features of certain distortion degree, and vice versa. Using this dictionary for image encoding with max pooling, we can obtain effective image representations. Finally, random forest is used to learn the mapping from feature space to subjective human scores. Random forest is an ensemble learning method that operates by constructing a multitude of decision trees at training time and predicting new data by averaging the predictions of all trees. Because this prediction process is closer to the process of subjective IQA, we adopt random forest instead of other models as the learning model.

Our contributions can be summarized as follows. First, we propose a novel NR-IQA framework using sparse filtering, one form of unsupervised feature learning, in combination with NSS information extracted from image patch. Additionally, we apply random forest to learn the regression model. In contrast, most existing NR-IQA methods use Support Vector Machine (SVM) with different kernels for regression. In our experiment, we find that random forest can achieve significantly better results than SVM.

The remainder of this paper is organized as follows. Section 2 describes details of the proposed framework. Experimental strategies and results are presented in Section 3. Finally, Section 4 concludes with a summary of our work.

2. FRAMEWORK FOR QAF

Figure 1 illustrates the pipeline of QAF. Key components in this framework include NSS-based local descriptor construction, quality-aware filter learning, local descriptor encoding and feature pooling, and regression. The details of these components are described in the following sections.

2.1. NSS-based local descriptor

In our approach, each image is represented by a set of local descriptors extracted from randomly sampled patches. With respect to local descriptor construction, we resort to two NSS-based features presented in [17], which are derived from the distribution of mean subtracted contrast normalized (MSCN) coefficients [21] and the distribution of products of pairs of adjacent MSCN coefficients [12, 17]. The main advantage of these two models over other NSS ones is that they do not require a mapping to a different domain such as wavelet and DCT, and thus make the feature extraction process computationally efficient.

The procedure of MSCN can be seen as a normalization process with respect to local brightness and contrast, which can be described as,

$$\bar{I}(x,y) = \frac{I(x,y) - \mu(x,y)}{\sigma(x,y) + \gamma} \tag{1}$$

where I is a given gray-scale image, x and y are spatial coordinates, $\gamma = 1$ is a constant that prevents instabilities from occurring when the denominator tends to zero and

$$\mu(x,y) = \sum_{\alpha=-K}^{K} \sum_{\beta=-L}^{L} w_{\alpha,\beta} I(x+\alpha, y+\beta) \tag{2}$$

$$\sigma(x,y) = \sqrt{\sum_{\alpha=-K}^{K} \sum_{\beta=-L}^{L} w_{\alpha,\beta} [I(x+\alpha, y+\beta) - \mu(x,y)]^2} \tag{3}$$

estimate the local mean and contrast, respectively, where $w = \{w_{\alpha,\beta} | \alpha = -K, ..., K, \beta = -L, ..., L\}$ is a 2D circularly-symmetric Gaussian weighting function sampled out to 3 standard deviations and rescaled to unit volume.

Despite MSCN, the sample distribution of the products of the pairs of adjacent MSCN coefficients computed along horizontal, vertical and diagonal orientations, $\bar{I}(x,y)\bar{I}(x,y+1)$, $\bar{I}(x,y)\bar{I}(x+1,y)$, $\bar{I}(x,y)\bar{I}(x+1,y+1)$ and $\bar{I}(x,y)\bar{I}(x+1,y-1)$, can also capture the degree of quality distortion.

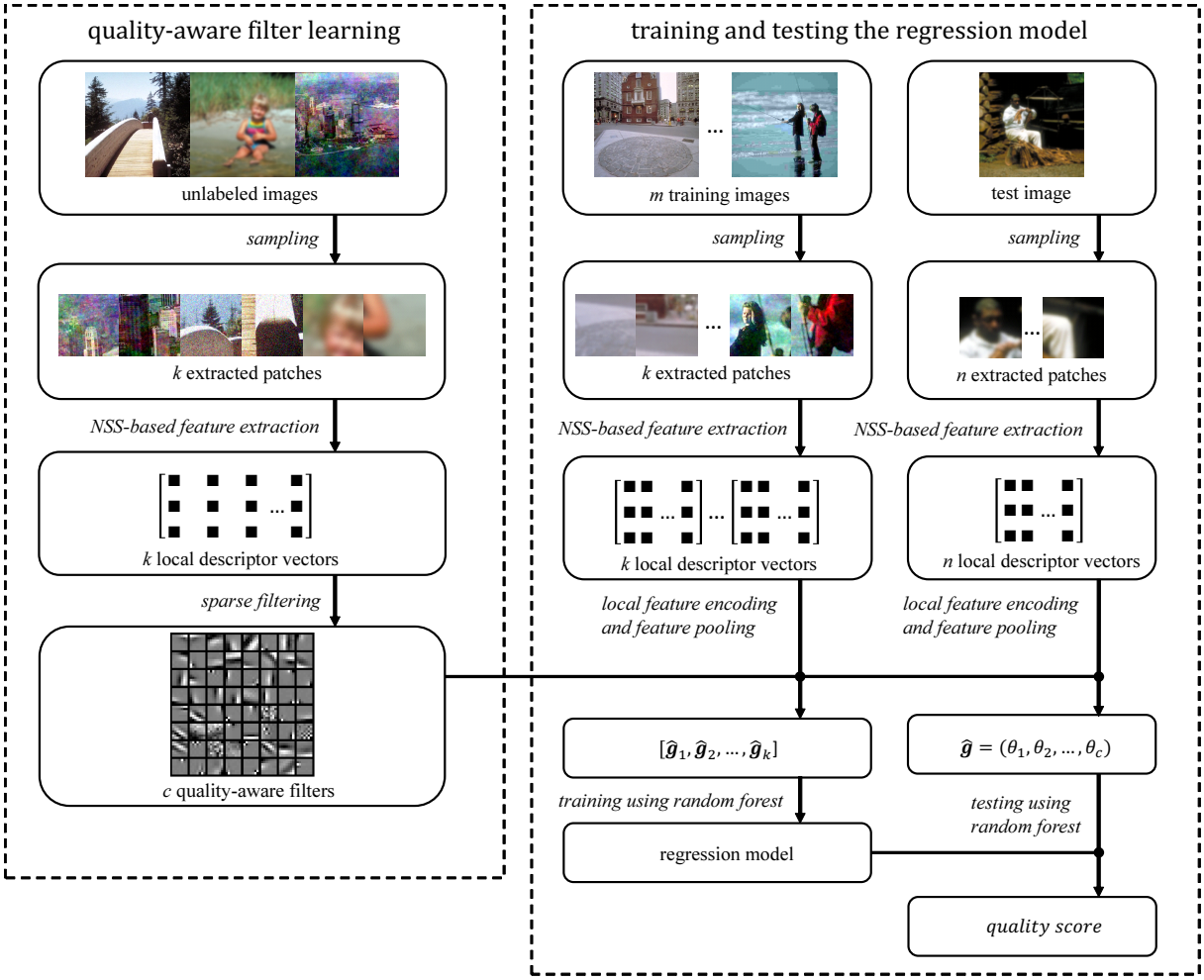

Fig. 1. Pipeline of the NR-IQA metric QAF.

Given a sampled image patch $P \in \mathcal{R}^{M \times N}$, where M and N are the patch height and width respectively, we can get a vector form as $\boldsymbol{f} = (p_1, p_2, ..., p_{M \times N \times 5})^T$ by applying the two introduced NSS model to it, and $\boldsymbol{f}$ is regarded as the descriptor of P. Consequently, each image can be represented as a set of descriptor vectors $\boldsymbol{F} = [\boldsymbol{f}_1, \boldsymbol{f}_2, ..., \boldsymbol{f}_n]$, where each $\boldsymbol{f}_i$ is a local descriptor vector and n is the number of patch sampled.

2.2. Quality-aware filter learning

Unspervised feature learning has been shown to be effective at learning representations that perform well on image, video and audio classification. In our apporach, we apply one simple yet efficient unsupervised feature method, namely sparse filtering [19], to a set of descriptor vectors, which are extracted from image patches of different quality degrees, to learn a quality-aware filter dictionary. This dictionary can map the patch-based local descriptors into good feature representations for the task of quality classification. The details of this process will be presented in the following paragraphs.

At first, we randomly sample image patches from a set of unlabeled training images which suffer from distortion at different degrees. Then, we compute from it a set of local descriptors: $\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_k]$, where $\boldsymbol{Y} \in \mathcal{R}^{d \times k}$, d is the dimension of each local descriptor and k is the number of patches. Then, we initialize a filter dictionary $\boldsymbol{X} \in \mathcal{R}^{v \times d}$ with random values, where v is the number of filters we aim to learn. Then, the objective of sparse filtering can be formulated as:

$$\widehat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}} \sum_{i=1}^{\Omega} \varphi |\boldsymbol{Z}| \tag{4}$$

where $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{Y}$, Ω is the number of entries in $\boldsymbol{Z}$, $|*|$ represents the soft-absolute operation $\boldsymbol{Z} = \sqrt{\varepsilon + \boldsymbol{Z}^2} (\varepsilon = 10^{-8})$ and φ stands for the operation of performing twice normalization on $|\boldsymbol{Z}|$ by rows and by columns using l_2-norm, respectively. An off-the-shelf L-BFGS [22] package can be used to optimize this objective function until convergence. The design rationale of sparse filtering is that the normalization op-

eration introduces competition, which makes some values in $\varphi|\boldsymbol{Z}|$ have to be large while the others are small(close to 0).

To have a more stable filter dictionary, the above process is repeated u times, and after we get a filter collection of $v \times u$ filters, we perform K-means on it to get $\boldsymbol{C}$ consisting of c filters. 256 randomly selected filters learned from sparse filtering are shown in Figure 2.

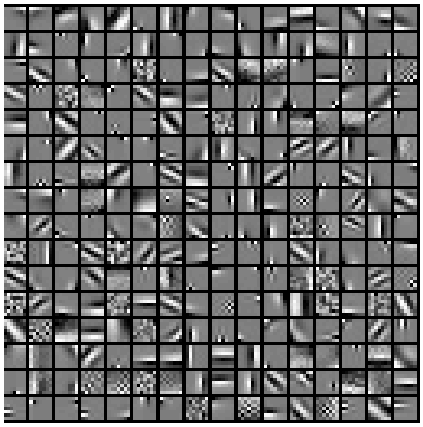

Fig. 2. A subset of quality-aware filters.

2.3. Local descriptor encoding and feature pooling

As described in 2.1, in our approach, each image is represented by a set of local descriptors $\boldsymbol{F} = [\boldsymbol{f}_1, \boldsymbol{f}_2, ..., \boldsymbol{f}_n]$, where $\boldsymbol{f}_i$ is a NSS-based local descriptor and n is the number of sampled patch. Once we learned the filter dictionary $\boldsymbol{C}$, we can encode $\boldsymbol{F}$ by first applying soft-absolute on $\boldsymbol{G} = \boldsymbol{CF}$, and then normalizing $\boldsymbol{G}$ by rows and by columns separately.

The feature encoding step provides us with a matrix $\boldsymbol{G}_{c\times n}$ for each image, where c is the number of filters in dictionary $\boldsymbol{C}$. In order to learn a regression model, we need a fixed-length feature vector. To achieve this purpose, we apply max pooling to the matrix $\boldsymbol{G}_{c\times n}$. Typically, there are two types of pooling strategies: average-pooling and max-pooling. In our experiment, we find that max-pooling can achieve better results. Specifically, for each column $\boldsymbol{g} = (\theta_1, \theta_2, ..., \theta_c)^T$ in $\boldsymbol{G}$, the max-pooling operated on $\boldsymbol{g}$ can be written as,

$$\theta_i = \begin{cases} 1, \theta_i = max(\theta_1, \theta_2, \theta_3, ..., \theta_c) \\ 0, else \end{cases} \tag{5}$$

Then, the image level feature can be obtained by summing up all the columns. This can be written as,

$$\widehat{\boldsymbol{g}} = sum(\boldsymbol{g}_1, \boldsymbol{g}_2, \boldsymbol{g}_3, ..., \boldsymbol{g}_n) \tag{6}$$

where $\widehat{\boldsymbol{g}} \in \mathcal{R}^c$.

2.4. Regression

Having a set of image representations and their corresponding subjective scores, we can treat NR-IQA as a regression problem. Specifically, in the training stage, these image representations and their corresponding scores are used to construct a regression model. In the test stage, a feature vector is extracted from the test image and then fed into the learned regression model to predict its quality score. Most existing NR-IQA approaches use SVM for its simplicity. However, as we all know, in subjective IQA experiments, the finally quality is obtained by averaging the evaluation from different people. Inspired by this, we adopt random forest [20], which consists of a multitude of trees, as the regression model in our experiments. We find that random forest can achieve significantly better result than SVM in our framework.

3. EXPERIMENTAL RESULTS AND DISCUSSION

3.1. Protocol

We evaluated the proposed metric QAF on two widely-used IQA datasets: LIVE [18] and CSIQ [23]. Both of them provide us with pristine images, distorted versions and associated subjective scores. Brief information of these two datasets is listed in Table 1.

Table 1. Benchmark image datasets used.

Dataset	Distorted Image No.	Distorted Type No.
LIVE	779	5
CSIQ	866	6

LIVE database consists of 29 reference images each with five different types of distortion at 5 to 6 levels. The types include JPEG2000, JPEG, white noise (WN), Gaussian blur (BLUR) and simulated fast fading Rayleigh channel(FF).

CSIQ database consists of 30 reference images and their degraded versions with six different types of distortion at 4 to 5 levels. The types include JPEG2000, JPEG, WN, BLUR, global contrast decrements, and Gaussian pink noise.

To evaluate the performance of QAF metric, two correlation coefficients between the prediction results and the subjective scores were adopted: the Spearman rank order correlation coefficient (SROCC), which is related to the prediction monotonicity, and the Pearson linear correlation coefficient (PLCC), which is related to the prediction linearity. A value close to 1 for SROCC and PLCC indicates a good performance for quality estimation. In all the experiment, we only reported the results on distorted images.

Five opinion-aware approaches, namely BIQI [7], BRISQUE [12], BLIINDS-II [11], DIIVINE [8], CORNIA [13] and two opinion-unaware approach, namely NIQE [17] and QAC [15], were used for comparison. Although NIQE and QAC did not require the process of training, to ensure a fair comparison across methods, we reported their result on test images only. The evaluation results on LIVE and CSIQ are summarized in Table 2 and 3.

Table 2. Performance Evaluation on LIVE.

Methods	80 %		50 %		10 %	
	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
BIQI	0.825	0.840	0.739	0.764	0.547	0.623
BRISUQE	0.933	0.931	0.917	0.919	0.806	0.816
BLIINDS-II	0.924	0.927	0.901	0.901	0.836	0.834
DIIVINE	0.884	0.893	0.858	0.866	0.695	0.701
CORNIA	0.940	0.944	0.933	0.934	0.893	0.894
NIQE	0.908	0.908	0.905	0.904	0.905	0.903
QAC	0.874	0.868	0.869	0.864	0.866	0.860
QAF	0.947	0.951	0.946	0.949	0.943	0.944

Table 3. Performance Evaluation on CSIQ.

Methods	80 %		50 %		10 %	
	SROCC	PLCC	SROCC	PLCC	SROCC	PLCC
BIQI	0.092	0.237	0.092	0.396	0.020	0.311
BRISUQE	0.775	0.817	0.736	0.781	0.545	0.596
BLIINDS-II	0.780	0.832	0.749	0.806	0.628	0.688
DIIVINE	0.757	0.795	0.652	0.716	0.441	0.492
CORNIA	0.714	0.781	0.678	0.754	0.638	0.732
NIQE	0.627	0.725	0.626	0.716	0.624	0.714
QAC	0.486	0.654	0.494	0.706	0.490	0.707
QAF	0.780	0.840	0.768	0.810	0.741	0.754

Table 4. Evaluation results when trained on LIVE and tested on CSIQ.

Method	SROCC	PLCC
BIQI	0.619	0.695
BRISQUE	0.557	0.742
BLIINDS2	0.577	0.724
DIIVINE	0.596	0.697
CORNIA	0.663	0.764
NIQE	0.627	0.716
QAC	0.490	0.708
QAF	0.701	0.715

3.2. Implementation details

The proposed framework contains a number of parameters that can be tuned. The following ones may have great impact on the result of our approach: 1) n: number of patches sampled from each image; 2) M and N: width and height of the raw patch; 3) $u \times v$: total number of filters learned through sparse filtering; 4) c: number of quality-aware filters; 5) $ntree$ and $mtry$: two parameters used in random forest.

In our experiment, we set n = 10000, M = N = 7, c=10000, $ntree$ = 1500, and $mtry$ = 250. There is no explicit constraints on the number $u \times v$. Empirically, it should not be small than 100000.

3.3. Performance evaluation on single database

In the current literature, most NR-IQA algorithms are only evaluated on LIVE IQA database using the experimental strategy mentioned in [10, 11, 13]. Specifically, for training-based algorithms, 23 reference images along with their distorted images were randomly selected for training, and the rest 6 reference images along with their degraded versions were used for testing. Such an experimental strategy mainly has two deficiencies. On one hand, LIVE database only contains 779 distorted images, so the size of test images is only about 150 (779*0.2), which is too samll. On the other hand, the fact of training on 80% images and testing on only 20% images is also weak. With the large size of the training set, it is very likely to have the problem of over-fitting, so we can not evaluate the generalization ability of the algorithms properly. Therefore, here we adopted another experimental strategy similar to the one proposed in [15]. Specifically, for training-based methods, we present their results under three settings: 80%, 50% and 10% randomly selected samples are used for training and the remainder are used for testing. The partition is randomly conducted 1000 times and we report the median result.

From Table 2 and Table 3, we can see that QAF outperforms its counterparts under different ratios of training samples. Additionally, although the performance of most existing NR-IQA methods decrease rapidly with the decrease of numbers of training samples, QAF seems to be robust to the number of training samples.

3.4. Cross-database evaluation

In this section, we performed a more comprehensive performance evaluation by training on LIVE dataset and testing on CSIQ datasets. For the five opinion-aware NR-IQA methods, their quality prediction models trained on the entire LIVE dataset are provided by the original authors. Thus, we directly use them for testing on CSIQ. The experimental result is presented in Table 4. We can see that QAF performs significantly better than all the other state-of-the-art NR-IQA algorithms.

4. CONCLUSIONS

In this paper, we proposed an effective blind image quality assessment method, namely QAF. The design rationale is to learn quality-aware filters by performing sparse filtering on the NSS-based extracted features. We use the learned filters to encode image and adopt random forest to learn the regression model. Extensive experiments validated that QAF yields much better quality prediction performance than all the competing methods and is more robust to the number of training samples.

5. ACKNOWLEDGEMENT

This work is supported by NSFC under grant no. 61201394, the Shanghai Pujiang Program under grant no. 13PJ1408700, and the Innovation Program of Shanghai Municipal Education Commission under grant no. 12ZZ029.

References

[1] Z. Wang and A.C. Bovik, *Modern Image Quality Assessment*, Morgan and Claypool Publishers, 2007.

[2] Z. Wang, A.C. Bovik, and B.L. Evans, "Blind measurement of blocking artifacts in images," in *ICIP*, 2000, pp. 981–984.

[3] H. Liu, N. Klomp, and I. Heynderickx, "A no-reference metric for perceived ringing artifacts in images," *IEEE Trans. IP*, vol. 20, no. 4, pp. 529–539, Apr. 2010.

[4] R. Ferzli and L.J. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB)," *IEEE Trans. IP*, vol. 18, no. 4, pp. 717–728, Apr. 2009.

[5] H.R. Sheikh, A.C. Bovik, and L.K. Cormack, "No-reference quality assessment using natural scene statistics: JPEG2000," *IEEE Trans. IP*, vol. 14, no. 11, pp. 1918–1927, Nov. 2005.

[6] Z. Wang, H.R. Sheikh, and A.C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," in *ICIP*, 2002, pp. 477–480.

[7] A. Mittal, R. Soundararajan, and A.C. Bovik, "A two-step framework for constructing blind image quality indices," *IEEE Signal Process. Letters*, vol. 17, no. 5, pp. 513–516, May 2010.

[8] A. Mittal, R. Soundararajan, and A.C. Bovik, "Blind image quality assessment: from natural scene statistics to perceptual quality," *IEEE Trans. IP*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.

[9] P. Ye and D. Doermann, "No-reference image quality assessment using visual codebooks," *IEEE Trans. IP*, vol. 21, no. 7, pp. 3129–3138, Jul. 2012.

[10] M.A. Sadd, A.C. Bovik, and C. Charrier, "A DCT statistics-based blind image quality index," *IEEE Signal Process. Letters*, vol. 17, no. 6, pp. 583–586, Jun. 2010.

[11] M.A. Sadd, A.C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the dct domain," *IEEE Trans. IP*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.

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

[13] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in *CVPR*, 2012, pp. 1098–1105.

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

[15] W. Xue, L. Zhang, and X. Mou, "Learning without human socres for blind image quality assessment," in *CVPR*, 2013, pp. 995–1002.

[16] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. IP*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.

[17] A.K. Moorthy and A.C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Process. Letters*, vol. 20, no. 3, pp. 209–212, Mar. 2013.

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

[19] J. Ngiam, P. Koh, Z. chen, S. Bhaskar, and A.Y.Ng, "Sparse filtering," in *NIPS*, 2011.

[20] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[21] J. Geusebroek and A. Smeulders, "The statistics of natural images," *Netw. Comput. Neural Syst.*, vol. 5, no. 4, pp. 517–548, 1994.

[22] M. Schmidt, "minfunc," 2005, http://www.cs.ubc.ca/ schmidtm/software/minFunc.html.

[23] E.C. Larson and D.M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy," *J. Electr. Imaging*, vol. 19, no. 1, pp. 001006:3440–3451, Mar. 2010.